

**THESE DE DOCTORAT DE
L'UNIVERSITE PIERRE ET MARIE CURIE**

Spécialité

MATHEMATIQUES

Ecole Doctorale de
Sciences Mathématiques de Paris Centre (ED386)

Présentée par

M. THOMAS ROMARY

Pour obtenir le grade de

DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE

Sujet de la thèse :

INVERSION DES MODELES STOCHASTIQUES DE MILIEUX HETEROGENES

soutenue le 19.12.2008 devant le jury composé de :

Jean Jacod, Université Paris VI

Eric Moulines, ENST

Fabien Campillo, INRIA

Lin-Ying Hu, IFP

Gilles Pagès, Université Paris VI

Christian Lantuéjoul, Centre de Géostatistique/ENSMP

Directeur de thèse

Rapporteur

Rapporteur

Examineur

Examineur

Examineur

Remerciements

Cette thèse a été réalisée au sein de la direction Ingénierie de Réservoir de l'IFP. Au terme de ces trois années, je souhaite adresser mes sincères remerciements :

A Monsieur Jean Jacod, qui m'a fait l'honneur de diriger ma thèse. Sur un sujet pourtant fort éloigné de ses domaines de prédilection, il aura fait preuve d'une remarquable disponibilité. Je tiens à lui exprimer toute ma gratitude pour ses conseils et remarques avisés, parfois teintés de pessimisme, son soutien et sa bienveillance, qui auront grandement contribué aux travaux réalisés durant ces trois ans.

A Monsieur Lin-Ying Hu, qui est à l'origine de cette thèse et a encadré mon travail à l'IFP. Je tiens à le remercier profondément pour la confiance et la liberté qu'il m'a accordées durant ces trois ans, pour les encouragements et conseils qu'il m'a prodigués lors de nos nombreuses discussions, enfin pour les qualités humaines dont il a fait preuve à mon égard.

A Monsieur Eric Moulines, qui m'a fait l'honneur d'être rapporteur de cette thèse. Je lui adresse mes plus vifs remerciements pour sa disponibilité, l'intérêt qu'il a porté à mon sujet de thèse et pour m'avoir fait découvrir les méthodes de chaînes de Markov en interaction.

A Monsieur Fabien Campillo, qui m'a fait l'honneur d'être rapporteur de ces travaux. Sa relecture minutieuse et ses remarques pertinentes auront grandement contribué à la rédaction finale de ce manuscrit.

A Messieurs Gilles Pagès et Christian Lantuéjoul qui m'ont fait l'honneur de bien vouloir juger mes travaux de thèse.

A tous les membres de la direction Ingénierie de Réservoir pour leur accueil chaleureux et la sympathie qu'ils m'ont témoignée tout au long de ces trois années. En particulier, je tiens à remercier Daniel et Frédéric pour nos parties de tennis acharnées, mes collègues fumeurs Alexandre, Christophe, Frédéric de nouveau, et Amir, mon collègue de bureau, pour leur bonne humeur ; plus largement, je souhaite remercier toutes les personnes qui ont contribué de près ou de loin à ce travail : Benoit Noetinger, Mickaele Le Ravalec, Frédéric Roggero, Patrick Lemonnier, Sylvie Hoguet, ainsi que tous ceux que j'oublie de citer mais qui se reconnaîtront sûrement et, j'en suis sûr, ne m'en tiendront pas rigueur.

A mes futurs collègues du centre de géostatistique de l'école des Mines de Paris, pour les remarques avisées qu'ils ont formulées lors des déjà nombreuses présentations de mon travail que j'ai effectuées à Fontainebleau.

A toute ma famille, pour leur soutien inconditionnel.

In fine, voglio ringraziare con tutto il mio cuore Valentina per essermi stata sempre vicina e per riempire, giorno dopo giorno, la mia vita di felicità. Questa tesi le è dedicata.

Résumé

La problématique du calage d'historique en ingénierie de réservoir, c'est-à-dire le calage des modèles géostatistiques aux données de production, est un problème inverse mal posé. Dans un cadre bayésien, sa résolution suppose l'inférence de la distribution de probabilité du modèle géostatistique conditionné aux données dynamiques, rendant compte à la fois de l'*a priori* géologique, exprimé dans le modèle géostatistique, et de l'adéquation aux données de production. Typiquement, la résolution de ce problème passe par la génération d'un ensemble de réalisations calées aux données, échantillon représentatif de cette distribution. Les modèles géostatistiques sont en général discrétisés sur des grilles de plusieurs centaines de milliers, voire des millions de blocs ; les approches classiques tentent de résoudre le problème inverse en considérant l'ensemble des blocs comme paramètres du modèle. Sa dimension est alors considérable et les méthodes d'échantillonnages deviennent impraticables sur un cas réel. Il convient alors de choisir une paramétrisation susceptible de réduire la dimension du problème. Dans la première partie de cette thèse, nous présentons une méthode de paramétrisation optimale des modèles géostatistiques basés sur les champs aléatoires gaussiens, à partir de leur décomposition de Karhunen-Loève (KL). Nous en décrivons les fondements théoriques, puis, sur des applications aux modèles de champs aléatoires gaussiens courants en géostatistique, selon des critères d'abord statistiques puis hydrodynamiques, nous quantifions la réduction de la dimension du problème offerte par cette paramétrisation. Dans la seconde partie, nous présentons les principes des méthodes de Monte-Carlo par Chaînes de Markov (MCMC) et les défauts des méthodes classiques pour la résolution du problème inverse dans le cadre bayésien. Nous développons alors l'approche par chaînes de Markov en interaction dont nous exposons les avantages. Enfin, les résultats obtenus par l'emploi conjoint de ces deux méthodes sont présentés dans deux articles. Une approche différente, passant par l'emploi de méthodes d'analyse de sensibilité, est également décrite dans un troisième article.

Mots-clés

problème inverse, géostatistique, décomposition de Karhunen-Loève, chaînes de Markov en interactions, calage d'historique, réduction de dimension, milieu poreux, statistique

HETEROGENEOUS MEDIA STOCHASTIC MODELS INVERSION

Abstract

The history matching problem in reservoir engineering, which consists in matching the geostatistical model to production data, is an ill-posed inverse problem. Its resolution implies to infer the probability distribution of the geostatistical model conditioned to the dynamical data, considering both the geological *a priori*, expressed in the geostatistical model, and the adequacy with production data. The aim is then to generate a representative sample from this distribution. Geostatistical models are usually discretized on a grid of hundreds of thousands even millions of blocks. Classical methods aim at solving the inverse problem considering the value of the random field at each gridblock as parameters. The huge dimension of this problem makes it intractable in practice. Therefore, we need a parameterization technique that will be able to reduce the problem dimension. The first part of this thesis is devoted to studying a parameterization method of Gaussian-related geostatistical models, based on their Karhunen-Loève (KL) expansion. We describe its theory and principles. Then, according to both statistical and dynamical (with respect to fluid flow results) criteria, we quantify the dimension reduction offered by this decomposition. In the second part, we present the principles of Markov chain Monte-Carlo (MCMC) methods. Particularly, we point out the defects of classical approaches when trying to solve a bayesian inversion problem. We then introduce the interacting Markov chains methods whose advantages are exposed. Finally, the results obtained by using jointly both methods are exhibited in two publications. A different approach, using sensitivity analysis methods, is also described in a third article.

Keywords

inverse problem, geostatistics, Karhunen-Loève expansion, interacting Markov chains, history matching, dimension reduction, porous media, statistics

Table des matières

Table des matières	9
Table des figures	12
1 Introduction	17
1.1 L'information disponible	17
1.1.1 Les données géologiques	18
1.1.2 Les données géophysiques	18
1.1.3 Les données dynamiques	19
1.1.4 Bilan	19
1.2 Modélisation géostatistique	19
1.2.1 Modélisation des propriétés pétrophysiques	20
1.2.2 Modélisation en faciès	21
1.2.3 Bilan	21
1.3 Le problème inverse	21
1.3.1 L'approche bayésienne	22
1.3.2 Le problème direct	23
1.3.3 Courte revue des méthodes de calage d'historique	24
1.4 Objectifs et résultats	26
1.4.1 Objectifs	26
1.4.2 Plan	26
1.4.3 Résultats et perspectives	27
2 Décomposition de Karhunen-Loève	31
2.1 Introduction	31
2.2 Quelques rappels sur les processus du second ordre	33
2.3 Théorème de Karhunen-Loève	36
2.3.1 Enoncé du théorème	36
2.3.2 Conservation de la variance	37
2.3.3 Le problème aux valeurs propres	38
2.3.4 Relation avec la représentation harmonique	41
2.4 Application aux modèles géostatistiques	43
2.4.1 Analyse statistique	43
2.4.2 Recherche d'un critère dynamique de choix du nombre de com- posantes	49
2.4.3 Discussion	55

2.5	Modèles de gaussiennes seuillées	56
2.5.1	Définitions	56
2.5.2	Choix du modèle de covariance pour la modélisation en faciès	57
2.5.3	Expériences numériques	60
2.5.4	Application aux écoulements	71
2.6	Autres applications de la décomposition de Karhunen-Loève	78
2.6.1	Karhunen-Loève et déformation graduelle	78
2.6.2	Karhunen-Loève comme méthode d'approximation	80
2.6.3	Approches alternatives	82
3	Méthodes McMC	85
3.1	Introduction	85
3.2	Motivations : calcul d'intégrales et de sommes	86
3.3	Chaînes de Markov	88
3.3.1	Notions de base	88
3.3.2	Propriétés	90
3.3.3	Mesure invariante	92
3.3.4	Convergence et théorèmes limites	94
3.3.5	La covariance dans les chaînes de Markov	96
3.4	L'algorithme de Metropolis-Hastings	96
3.4.1	Définition	97
3.4.2	Propriétés de convergence	98
3.4.3	Echantillonneur indépendant	99
3.4.4	Marches aléatoires	100
3.4.5	Echantillonneurs autorégressifs	101
3.4.6	Echantillonneur de Langevin	101
3.4.7	Schémas adaptatifs	102
3.4.8	Recuit et trempe simulés	106
3.4.9	Considérations pratiques	108
3.5	Application	109
3.5.1	Le problème inverse	109
3.5.2	Comparaison des performances des algorithmes MH	110
3.6	Chaînes en interaction	115
3.6.1	Introduction	115
3.6.2	Principe général	116
3.6.3	Description des algorithmes	117
3.6.4	Résultats de convergence	120
3.6.5	Considérations pratiques	122
3.6.6	Application	124
4	Applications	127
4.1	Integrating production data under uncertainty by parallel interacting Markov chains on a reduced dimensional space	127
4.2	History matching of truncated Gaussian models by parallel interacting Markov chains on a reduced dimensional space	149

4.3 Reservoir forecasting under uncertainty	159
Bibliographie	175

Table des figures

1.1	<i>Des données statiques au modèle géologique.</i>	22
2.1	<i>Variogrammes considérés.</i>	44
2.2	<i>Somme cumulée des valeurs propres, cas 1D, $pp = H$.</i>	46
2.3	<i>Somme cumulée des valeurs propres, cas 1D, $pp = H/4$.</i>	46
2.4	<i>Somme cumulée des valeurs propres, cas 2D, $pp = H$.</i>	46
2.5	<i>Somme cumulée des valeurs propres, cas 2D, $pp = H/4$.</i>	46
2.6	<i>Somme cumulée des valeurs propres, cas 3D, $pp = H$.</i>	46
2.7	<i>Somme cumulée des valeurs propres, cas 3D, $pp = H/4$.</i>	46
2.8	<i>Somme cumulée des valeurs propres, cas 1D, effet pépité $p=0.1$, $pp = H$.</i>	48
2.9	<i>Somme cumulée des valeurs propres, cas 2D, effet pépité $p=0.1$, $pp = H$.</i>	48
2.10	<i>Somme cumulée des valeurs propres, cas 3D, effet pépité $p=0.1$, $pp = H$.</i>	48
2.11	<i>Lignes de courant reliant un puits injecteur et un puits producteur sur une carte de perméabilité.</i>	50
2.12	<i>water cut des 100 simulations de référence et variance de l'échantillon (en trait discontinu rouge).</i>	52
2.13	<i>Watercut obtenu pour la réalisation complète et à différents ordres de troncature.</i>	53
2.14	<i>Erreur relative sur le water cut en fonction du nombre de composantes de l'approximation.</i>	53
2.15	<i>Erreur absolue (en jours) commise vis-à-vis du temps de percée en fonction du nombre de composantes.</i>	53
2.16	<i>Erreur relative moyenne sur le water cut en fonction du nombre de composantes de l'approximation et intervalle de confiance à 95% associé.</i>	54
2.17	<i>Erreur absolue moyenne (en jours) commise vis-à-vis du temps de percée en fonction du nombre de composantes et intervalle de confiance à 95% associé.</i>	54
2.18	<i>Comportement d'une trajectoire brownienne (pas de discrétisation de 0.0001) lors du passage d'un seuil quelconque.</i>	58
2.19	<i>Représentation du variogramme de Matérn pour ν variant de 0 à 65 selon un pas de discrétisation de 1/8 et portée pratique correspondante calculée et estimée.</i>	60
2.20	<i>Exemple d'approximation d'un champ 2D en facies à différents ordres de troncature.</i>	63
2.21	<i>Variogrammes considérés et variogrammes du modèle en facies (seuil 0.1) correspondant.</i>	63
2.22	<i>Taux de reconstitution erronée, cas 1D, $pp = H$.</i>	64

2.23	Taux de reconstitution erronée, cas 1D, $pp = H/4$	64
2.24	Taux de reconstitution erronée, cas 2D, $pp = H$	64
2.25	Taux de reconstitution erronée, cas 2D, $pp = H/4$	64
2.26	Taux de reconstitution erronée, cas 3D, $pp = H$	64
2.27	Taux de reconstitution erronée, cas 3D, $pp = H/4$	64
2.28	Histogrammes réalisés sur un échantillon de taille 10000 pour les 10 premières composantes de la décomposition du modèle en faciès (dimension 1) et densité gaussienne centrée réduite (en rouge).	66
2.29	Histogrammes des 100 premières variables de la décomposition du modèle en faciès pour le variogramme de Matérn, $\nu = 2$ et une portée pratique $pp = H/4$	69
2.30	Réalisations du modèle en faciès approximé pour le variogramme de Matérn, $\nu = 2$ et une portée pratique $pp = H/4$. A gauche, les réalisations obtenues en tronquant une réalisations en faciès, à droite obtenues en simulant selon l'hypothèse de v.a. gaussiennes indépendantes	70
2.31	Erreur relative sur le water cut en fonction du nombre de composantes de l'approximation (réalisation 1).	72
2.32	Erreur absolue (en jours) commise vis-à-vis du temps de percée en fonction du nombre de composantes (réalisation 1).	72
2.33	Erreur relative sur le water cut en fonction du nombre de composantes de l'approximation (réalisation 67).	73
2.34	Erreur absolue (en jours) commise vis-à-vis du temps de percée en fonction du nombre de composantes (réalisation 67).	73
2.35	Cartes de faciès correspondant à différents nombres de composantes dans la décomposition (réalisation 67).	74
2.36	Courbes de water cut correspondant à différents nombres de composantes dans la décomposition (réalisation 67).	75
2.37	Erreur relative moyenne sur le water cut en fonction du nombre de composantes de l'approximation et intervalle de confiance à 95% associé.	76
2.38	Erreur absolue moyenne (en jours) commise vis-à-vis du temps de percée en fonction du nombre de composantes et intervalle de confiance à 95% associé.	76
2.39	Taux de reconstitution erronée pour le cas étudié.	77
3.1	Densité a posteriori.	110
3.2	KDE du posterior construit à partir d'un échantillon i.i.d. de taille 10000.	111
3.3	KDE du posterior construit à partir de la chaîne générée par l'algorithme IMH.	112
3.4	KDE du posterior construit à partir de la chaîne générée par l'algorithme SIMH.	112
3.5	KDE du posterior construit à partir de la chaîne générée par l'algorithme LMH.	113
3.6	KDE du posterior construit à partir de la chaîne générée par l'algorithme ASIMH	113
3.7	Evolution des paramètres adaptées au cours de l'algorithme ASIMH.	114

3.8	Taux d'acceptation empirique le long de la chaîne pour chacun des quatre algorithmes, échelle logarithmique.	115
3.9	Autocorrélations le long de la chaîne pour chacun des quatre algorithmes, échelle logarithmique.	115
3.10	Représentation schématique d'un échange.	119
3.11	Autocorrélations le long de la chaîne pour l'algorithme PT, échelle logarithmique.	125
3.12	Autocorrélations le long de la chaîne pour l'algorithme EES, échelle logarithmique.	125
3.13	Autocorrélations le long de la chaîne pour l'algorithme PIR, échelle logarithmique.	125
3.14	KDE du posterior construit à partir des résultats de l'algorithme PT. .	126
3.15	KDE du posterior construit à partir des résultats de l'algorithme EES. .	126
3.16	KDE du posterior construit à partir des résultats de l'algorithme PIR. .	126

Chapitre 1

Introduction

L'étude des réservoirs d'hydrocarbures a pour objectif l'optimisation du rendement économique de l'extraction du pétrole (et/ou du gaz). Dans le contexte actuel d'épuisement des gisements en exploitation et de demande toujours croissante en hydrocarbures, le besoin d'optimiser la production de pétrole brut se fait toujours plus pressant. Selon les types de gisements et les techniques de récupération employées, on peut extraire de 10% à 70% du volume global d'hydrocarbure contenu dans le réservoir. Les méthodes et stratégies développées par l'ingénierie de réservoir permettent d'accroître sensiblement ce rendement. Plus précisément, l'ingénierie de réservoir est la discipline dédiée à l'étude de réservoir par l'intégration de données hétéroclites, pour accroître notre connaissance quant à la structure géologique du gisement ; à la simulation des déplacements de fluide en leur sein ; à l'imagination et la modélisation de techniques innovantes de récupération ; enfin, à l'estimation de la faisabilité et la modélisation physique du stockage de CO_2 , dans les réservoirs déplétés. Pratiquement, il s'agit d'intégrer l'ensemble des données disponibles, qui sont de natures diverses (géologiques, géophysiques et dynamiques), dans un modèle de réservoir qui permettra de prédire la production et d'optimiser la stratégie d'exploitation, ou d'évaluer le risque de fuite à très long terme, dans le cas du stockage de CO_2 .

Dans cette introduction, nous allons nous attacher dans un premier temps à décrire les différents types de données disponibles en section 1.1 ; puis nous passerons en revue les différents modèles stochastiques de milieux hétérogènes, en section 1.2. Enfin, en section 1.3, nous décrirons plus spécifiquement le problème de l'intégration des données dynamiques avant de décrire les objectifs de la thèse, le plan du manuscrit et les principaux résultats obtenus en section 1.4.

1.1 L'information disponible

L'information dont dispose l'exploitant d'un réservoir d'hydrocarbures se divise en trois principaux types :

1. les données géologiques,
2. les données géophysiques,
3. les données dynamiques.

1.1.1 Les données géologiques

Les hydrocarbures, qui se trouvent dans le réservoir sous forme liquide et/ou gazeuse, proviennent de la dégradation de matières organiques présentes dans les sédiments, sous des conditions de température et de pression adéquates, dans un domaine géologique limité par des barrières imperméables. Une modélisation du bassin sédimentaire, à une échelle plus large que celle du réservoir, est donc un préalable à la caractérisation du réservoir. Elle est basée sur la connaissance des processus géologiques de sédimentation et de tectonique au cours des âges ainsi que sur les différentes données disponibles.

Les données géologiques peuvent être locales ou globales. Dans la première catégorie, on placera les données de carottage, qui renseignent directement sur la géologie locale lors du forage d'un puits ou d'un simple sondage. On trouve également dans cette catégorie les données issues des techniques de diagraphies qui fournissent des mesures des propriétés pétrophysiques le long d'un puits à l'aide de sondes. Ces données sont indirectes dans la mesure où elles proviennent d'appareils de mesure et peuvent être entachées d'une erreur. Plus précisément, ces deux types de données apportent une information sur le type de roche sédimentaire rencontrée, *e.g.* grès, argile, carbonate ou évaporite, que l'on dénommera *lithofaciès* ou plus simplement *faciès*, ses propriétés pétrophysiques (perméabilité, porosité) et les saturations locales des différents fluides. Il convient de conditionner localement le modèle de réservoir à ces données.

Les données géologiques globales viennent principalement d'observations d'affleurement, c'est-à-dire de pans entiers de roche de même type que celui que l'on trouve dans le réservoir, rendus visibles au gré des mouvements tectoniques et qui fournissent une information de type conceptuel.

L'ensemble de ces données géologiques, via leur interprétation par un géologue, peut être intégré dans le modèle de bassin sédimentaire, puis dans le modèle de réservoir. Ce modèle renseignera sur la compréhension des environnements de dépôts des différentes unités du réservoir, les distributions de perméabilité et de porosité associées aux différents faciès du réservoir, ainsi que la géométrie du modèle de réservoir sur lequel sera conduite la simulation d'écoulement.

1.1.2 Les données géophysiques

Les données géophysiques sont issues de l'emploi de méthodes sismiques. La sismique consiste à exploiter les informations générées par la propagation d'ondes de compression et de cisaillement dans le sous-sol. Ces ondes se réfléchissent partiellement lorsqu'elles rencontrent un fort contraste d'*impédance* (produit de la densité par la vitesse sismique). En exploitant les temps de parcours sources-récepteurs ainsi que l'amplitude du champ d'onde, on est capable de produire une image structurale des couches géologiques du sous-sol. Cette dernière demeure imprécise en raison du fort niveau de bruit affectant les données. Elle permet néanmoins de déterminer précisément la géométrie du domaine du réservoir et notamment la disposition des principales failles. Elle fournit également des informations sur les distributions de perméabilité et de porosité. Enfin, elle donne également une idée de la position initiale des fronts de saturation en huile et en eau.

Différentes campagnes sismiques peuvent être menées au cours de l'exploitation d'un gisement, elles permettent notamment de mettre à jour les données de saturation.

1.1.3 Les données dynamiques

Les données dynamiques sont disponibles une fois initiée la phase d'exploitation du réservoir. Elles correspondent à des données de production : données de pression et d'écoulement de fluides au niveau des puits. Ces données proviennent de capteurs et instruments de production (têtes de puits, séparateur eau/huile, etc.) et sont généralement entachées d'erreurs de mesures. En outre, les expérimentations menées en laboratoire sur les échantillons d'huile extraite permettent de déterminer les données *PVT* (pression, volume, température) desquelles on peut déduire des grandeurs portant sur la qualité de l'huile, telles que la compressibilité ou la viscosité, nécessaires à la paramétrisation de la simulation d'écoulement.

1.1.4 Bilan

Les données géologiques et sismiques vont fournir une information que l'on qualifiera de *statique*, puisqu'elles sont disponibles avant la mise en exploitation. Elles permettront de construire un modèle structural du réservoir, dont le modèle dit stratigraphique est une version discrétisée sur un maillage cartésien, et de définir un modèle *géostatistique* ([Chiles and Delfiner, 1999]) de la distribution spatiale des différents faciès et de leur propriétés pétrophysiques, lui aussi discrétisé sur la grille. Des réalisations (ou simulations) de ce modèle pourront être générées afin de déterminer par exemple une estimation du volume global d'hydrocarbures ainsi que la quantité récupérable. Enfin, ces données permettront de déterminer les conditions initiales de pression et de saturation indispensables à la conduite de la simulation d'écoulement.

Les données dynamiques, en revanche, apportent une information indirecte sur la structure du réservoir et ses propriétés pétrophysiques. On cherchera donc à *conditionner* le modèle géostatistique aux données de production au fur et à mesure de l'exploitation du réservoir. On parlera de *calage d'historique* (*history matching*) lorsque l'on conditionnera le modèle géostatistique aux données de production. Avant d'aborder ce problème, nous allons décrire brièvement les différents modèles géostatistiques de réservoir.

1.2 Modélisation géostatistique

La géostatistique est une discipline qui fit son apparition au début des années 60, sous l'impulsion de Georges Matheron [Matheron, 1965]. Dans un premier temps cantonnée aux applications minières, son utilisation s'est étendue au domaine pétrolier dans les années 70. Elle couvre aujourd'hui un large champ d'application allant de la météorologie aux sciences médicales en passant par l'écologie.

Les modèles géostatistiques appliqués à l'ingénierie de réservoir, permettent de simuler la répartition spatiale des différents faciès dans le réservoir et de leurs propriétés pétrophysiques. On distingue par conséquent deux types de modélisation :

- la modélisation des grandeurs continues.
- la modélisation en faciès.

1.2.1 Modélisation des propriétés pétrophysiques

La modélisation des grandeurs continues comme la perméabilité ou la porosité repose principalement sur l'emploi de *fonctions aléatoires* ou *champs aléatoires* d'ordre 2. Les modèles employés en ingénierie de réservoir sont généralement liés à des modèles gaussiens dont les paramètres (espérance et covariance) sont estimés à partir des données statiques. Une éventuelle transformation, appelée *anamorphose*, permet de se ramener à un modèle gaussien.

En géostatistique, plutôt que la covariance on préfère la notion plus générale de variogramme. Dans le cas des modèles de second ordre, le variogramme se déduit directement de la covariance. Ce paramètre du champ aléatoire permet d'en déterminer les caractéristiques spatiales. En particulier, on peut définir la *portée pratique* qui mesure la *longueur de corrélation* le long d'un axe. Autrement dit, il s'agit de la distance au-delà de laquelle la corrélation entre les valeurs du champ est inférieure à 0.95, dans la direction de l'axe considéré. On définit la direction principale comme la direction selon laquelle la portée est maximale. La plupart des modèles employés en pratique sont anisotropes, *i.e* la portée dans la direction orthogonale à la direction principale est strictement inférieure à la portée le long de la direction principale. Par exemple, la portée prend la forme d'un ellipse dans le cas d'un champ bidimensionnel. Il existe d'autres types d'anisotropie, voir [Chiles and Delfiner, 1999] pour plus de détail. Le second facteur du variogramme qui va influencer sur les caractéristiques spatiales du champ est son comportement à l'origine : linéaire, il va impliquer un champ non-différentiable, deux fois différentiable, le champ associé sera différentiable. Nous approfondissons ce point dans le paragraphe 2.5.2. Le variogramme est donc un paramètre extrêmement important dans l'emploi des modèles de fonction aléatoire. Il importe donc que le modèle employé soit choisi avec précaution (voir [Chiles and Delfiner, 1999] pour un exposé des principes de la variographie).

L'espérance et le variogramme de la fonction aléatoire peuvent être ou non stationnaires. En pratique, on considère un jeu de paramètres (espérance et variogramme) par couche géologique, considérées indépendantes.

La pertinence de l'emploi des modèles de fonctions aléatoires gaussiennes a été remise en cause à plusieurs reprises et des modélisations alternatives ont été proposées. On peut citer principalement [Roislien and Omre, 2006] qui propose un modèle de fonctions aléatoires basées sur une généralisation de la loi de Student (*t-fields*), et [Gunning, 2002] qui propose l'emploi de champs aléatoires Lévy-stables pour simuler les hétérogénéités. Cependant, les difficultés pratiques liées à l'emploi de tels modèles — la simulation des champs de Levy pose des problèmes pratiques sur de grandes grilles ([Gunning, 2002]), et celle des *t-fields*, obtenus par sommation de variables gaussiennes, implique un nombre considérable de variables aléatoires — les disqualifient pour l'utilisation qui en est faite en ingénierie de réservoir, notamment dans la problématique du calage d'historique.

1.2.2 Modélisation en faciès

Deux principales modélisations existent pour simuler la répartition spatiale des faciès, par nature discontinue : ceux dérivés des modèles de fonctions aléatoires gaussiennes comme les gaussiennes seuillées et les pluri-gaussiennes (cf. *e.g.* [Chiles and Delfiner, 1999]), et les modèles dits multi-points ([Strebel, 2002]).

Les premiers consistent à définir une fonction de seuillage, qui, appliquée à une ou plusieurs fonctions aléatoires continues, corrélées ou non, permet de faire apparaître les différents faciès en respectant leurs proportions, préalablement renseignées.

Les seconds, proches des modèles de reproduction de textures en analyse d'image, consistent à reproduire les motifs d'une image d'apprentissage. Cette image d'apprentissage provient soit d'une expertise géologique, soit d'un modèle de réservoir, dit *génétique*, plus fin que le modèle de simulation de bassin, simulant les dépôts sédimentaires ayant eu lieu au cours des âges par des méthodes stochastiques (cf. [Jacod and Joathon, 1971] et [Lopez, 2003]). Pour ces modèles, l'impact de l'image d'apprentissage sur l'ensemble des simulations que l'on peut effectuer par la méthode multi-point n'est pas clair.

Enfin, notons l'existence des modèles dits *booléens* ou *objet* ([Chiles and Delfiner, 1999]) dont le principe consiste à simuler la position et la forme d'objets géologiques prédéfinis comme des lobes ou des chenaux.

1.2.3 Bilan

Quel que soit le type de modèle considéré, en faciès ou continu, il procure une information *a priori*, intégrant dans sa paramétrisation une part des données statiques. Il peut être ensuite conditionné relativement facilement aux données locales (carottes). Pour illustrer, nous présentons en figure 1.1 une schématisation des étapes de la construction du modèle géologique. Il s'agit alors d'intégrer les données dynamiques en conditionnant ce modèle, problème que nous décrivons dans la section suivante.

1.3 Le problème inverse

Dans la problématique du calage d'historique, l'ingénieur de réservoir observe une donnée de production

$$d = F(X), \tag{1.1}$$

fonction de la configuration du réservoir, par exemple la répartition spatiale de sa perméabilité, notée X , et où la fonction F représente l'écoulement des fluides au sein du milieu poreux. Etant donnée cette observation, il souhaite caractériser le réservoir, c'est-à-dire identifier la véritable configuration géologique du réservoir, pour en déduire la production future et optimiser la stratégie d'exploitation. Pour ce faire, il doit réaliser une inversion de l'équation 1.1. C'est un problème inverse mal-posé : la fonction F n'est *absolument pas* bijective et il n'y a donc pas unicité de la solution ; la dimension de X est de l'ordre de 10^4 à 10^7 tandis que celle de d est de l'ordre de 10 à 100.

En outre, l'ingénieur dispose d'outils imparfaits : la fonction d'écoulement F est approximée par un simulateur, décrit au paragraphe 1.3.2, et les outils de mesures de la

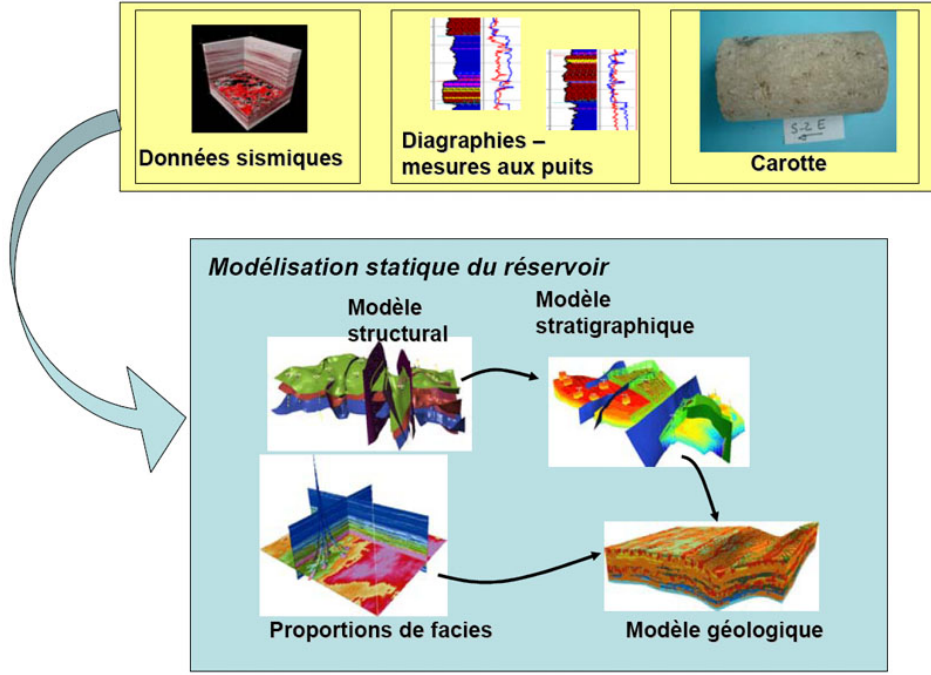


FIG. 1.1 – Des données statiques au modèle géologique.

donnée d entraînent une erreur sur celle-ci. Il est également nécessaire de tenir compte de l'information *a priori* que l'on possède sur le milieu poreux sous la forme du modèle géostatistique discrétisé dont on notera la loi $\mathbb{P}(X)$ et des valeurs en certains sommets de la grille, issues des données géologiques locales, qui le conditionnent. Ainsi, il ne s'agit pas tant de résoudre un problème d'inversion que de conditionner le champ aléatoire à la donnée de $d = F(X)$, elle même entâchée d'une erreur. Le cadre théorique adapté à ce problème est celui de l'inversion bayésienne, défini *e.g.* dans [Tarantola, 2005], et décrit dans le paragraphe suivant. F désigne désormais le simulateur d'écoulement.

1.3.1 L'approche bayésienne

On considère la loi *a priori* du milieu poreux discrétisé, c'est-à-dire défini sur une grille, $\mathbb{P}(X)$. On définit $\mathbb{P}(d|X)$ la loi modélisant l'erreur de mesure de d et l'erreur d'approximation faite par le simulateur d'écoulement : $(d - F(X))$, et la loi jointe $\mathbb{P}(d, X)$. En supposant que chacune de ces lois admet une densité, non triviale par rapport à la mesure de Lebesgue, notée $f(\cdot)$, la densité conditionnelle de X sachant d , via le théorème de Bayes, prend la forme suivante :

$$f(X|d) = \frac{f(d|X)f(X)}{\int f(d, X)dX}. \quad (1.2)$$

Il s'agit donc d'inférer la loi $\mathbb{P}(X|d)$, c'est-à-dire la loi du milieu discrétisé X conditionné par la donnée de production d , autrement dit la loi *a posteriori*. Dans le problème de calage d'historique, si l'on considère un *a priori* gaussien sur X ainsi que sur $(d|X)$, d'espérances et variances respectives μ et Γ , et $F(X)$ et C_d , la densité *a posteriori* (1.2)

1.3. LE PROBLÈME INVERSE

prend alors la forme suivante :

$$f(X|d) \propto e^{\left(-\frac{1}{2}\|d-F(X)\|_{C_d^{-1}}^2 - \frac{1}{2}\|X-\mu\|_{\Gamma^{-1}}^2\right)}, \quad (1.3)$$

où le symbole \propto signifie proportionnel à. La constante de normalisation est indépendante de X mais dépend de d . Dans le cas où X est un modèle de gaussienne seuillée, dont on notera S la fonction de seuillage, il suffit de remplacer F par $G = F \circ S$ dans (1.3).

La constante de normalisation dans l'expression ne peut être évaluée en raison de la dimension de X : son estimation demanderait un nombre incommensurable d'évaluations du problème direct. Il s'agira donc de construire un échantillon représentatif de (1.3).

Avant d'exposer les enjeux de cette approche et les méthodes existantes, nous présentons une brève description du problème direct, qui vise à calculer $F(X)$ quand on se donne une réalisation de X .

1.3.2 Le problème direct

Les écoulements en milieu poreux sont un phénomène complexe à modéliser car ils dépendent d'un grand nombre de paramètres. Ils dépendent non seulement des propriétés géométriques et pétrophysiques du milieu, mais également des conditions de pression et de température du réservoir, de la qualité de l'huile et des interactions entre les différents composants (eau, huile, gaz). Il existe en outre une forte incertitude sur les paramètres régissant les conditions aux limites. On dénombre six principaux types de modélisation plus ou moins simplifiés, liés aux techniques de récupération employées (injection de polymères, de vapeur pour l'extraction des huiles lourdes,...) ou aux types de réservoir (fissuré ou non). Les modèles d'écoulement intègrent donc différentes lois physiques : loi de conservation de la masse, loi de Darcy, les équations d'état des fluides (ou phases) ainsi que les conditions d'équilibre thermodynamique. Nous décrivons ici les deux relations fondamentales : la loi de conservation de la masse et la loi de Darcy. La loi de conservation de la masse signifie que la variation de la masse d'une phase dans une unité de volume du milieu poreux est égale à la différence des quantités entrantes et sortantes sur une unité de temps. Pour chacune des phases p , elle s'écrit :

$$\frac{\partial}{\partial t} (\rho_p S_p \phi) + \nabla (U_p \rho_p) = q_p, \quad (1.4)$$

où q_p désigne le débit d'injection ou de production de la phase p , ρ_p et S_p sa densité et sa saturation, ϕ la porosité, et U_p la vitesse d'écoulement de la phase.

Outre l'équation de conservation de la masse, une relation liant la vitesse d'écoulement U_p au gradient de pression P_p , est donnée par la loi de Darcy :

$$U_p = -\frac{k_{r,p}}{\mu_p} \mathbb{K} \nabla (P_p - \rho_p g z), \quad (1.5)$$

où g représente la gravité, z la profondeur, \mathbb{K} le tenseur de perméabilité de la roche, $k_{r,p}$ la perméabilité relative de la phase p , μ_p sa viscosité et P_p la pression de la phase

p .

La perméabilité relative $k_{r,p}$ quantifie la capacité de la phase p à circuler dans la roche en présence des autres phases. Elle est fonction des saturations des différentes phases dans le volume poreux.

Les inconnues de ce modèle sont ainsi la pression de chaque phase P_p et sa saturation S_p , dont les conditions initiales doivent être définies à $t = 0$.

Ces équations sont résolues numériquement par la méthode des différences finies. Un facteur important pour la modélisation d'écoulement est la taille du maillage. Il est préférable de travailler sur une grille de simulation aussi fine que possible afin de capturer les hétérogénéités du réservoir. Cependant, pour des raisons pratiques évidentes, il convient de faire un compromis entre la finesse de modélisation et les capacités de calcul. Ainsi, les simulations d'écoulement sont généralement effectuées sur un maillage relativement grossier faisant appel à des techniques de mise à l'échelle pour calculer les propriétés équivalentes du réservoir sur les mailles grossières. En effet, la construction du modèle statique est réalisée sur un maillage fin, dont le pas de discrétisation varie de 10 à 100 mètres horizontalement et de 0.1 à 1 mètre verticalement, comprenant jusqu'à 100 millions de mailles, alors que la modélisation de l'écoulement est effectuée sur un maillage d'échelle de l'ordre de la centaine de mètres horizontalement et de quelques mètres verticalement, et comprenant de 10 000 à 100 000 mailles. Lors de ce changement d'échelle, de l'information est agrégée, ce qui contribue à l'incertitude sur la caractérisation des propriétés du réservoir. Cependant, même en réduisant le nombre de mailles du modèle, le temps de calcul de l'écoulement reste élevé et est un des principaux écueils des méthodes d'assimilation des données dynamiques.

Notons enfin l'existence des modèles d'écoulement dits à *lignes de courants* (*stream-line*) où le problème tridimensionnel (1.4) est ramené à une collection de problèmes unidimensionnels le long de lignes de courants déterminées en fonction du gradient de pression et des vitesses d'écoulement. Ces modèles procurent généralement un gain de temps de calcul de l'ordre de 1/3 ([Oliver et al., 2001]). Leurs performances en terme d'erreur de calcul par rapport aux modèles aux différences finies classiques restent controversées.

Pour une réalisation donnée du modèle géostatistique du réservoir, le simulateur d'écoulement générera des données de production simulées que l'on pourra alors comparer aux données observées. Différentes méthodologies peuvent être mises en place pour résoudre le problème du calage d'historique, nous en décrivons quelques-unes dans le paragraphe suivant.

1.3.3 Courte revue des méthodes de calage d'historique

La description de l'approche bayésienne pour la problématique du calage d'historique remonte aux années 70 ([Gavalas et al., 1976]). Cependant, pendant longtemps, la plupart des méthodes de calage d'historique consistaient à déterminer le maximum *a posteriori* de (1.3), principalement en raison de ressources informatiques limitées. Le problème se traduit ainsi en un problème d'optimisation de la fonction objectif :

$$O(X) = \frac{1}{2} \|d - F(X)\|_{C_d^{-1}}^2 + \frac{1}{2} \|X - \mu\|_{\Gamma^{-1}}^2, \quad (1.6)$$

avec les notations du paragraphe 1.3.1. Dans ce cadre, l'effort a principalement porté sur la diminution de la dimension de l'espace dans lequel on effectue l'optimisation déterministe (descente de gradient), par des méthodes de paramétrisation. En effet, la dimension du problème (1.6) est celle de X et correspond au nombre de mailles de la grille sur laquelle X est discrétisé ; elle varie en général de dix mille à plusieurs millions de mailles. Dans [Gavalas et al., 1976], les auteurs considèrent l'emploi de la décomposition de Karhunen-Loève ([Loève, 1955]) pour réduire le nombre de composantes, nous approfondirons ce point dans le chapitre 2. Dans [Marsily et al., 1984], les auteurs proposent d'effectuer l'optimisation sur un certain nombre de points du champ aléatoire, les autres valeurs du champ étant interpolées par krigeage (cf. [Chiles and Delfiner, 1999]). On trouve également la méthode de déformation graduelle [Hu, 2000], qui permet d'effectuer l'optimisation dans un sous-espace généré par une combinaison d'un ensemble de réalisations du champ aléatoire ; les paramètres à optimiser sont alors ceux de cette combinaison. Cette méthode est employée à l'IFP pour les applications industrielles, on la décrit plus précisément dans le paragraphe 2.6.1. Une méthode équivalente a été développée à Stanford ([Caers and Hoffman, 2006]).

Pour une utilisation pratique de ces méthodes dans une optimisation, il est nécessaire de savoir calculer les gradients du simulateur d'écoulement dans la direction des paramètres d'intérêt. Ils peuvent être calculés soit numériquement, en effectuant un certain nombre de simulations additionnelles dans un voisinage de la réalisation considérée, soit en résolvant le système d'équation adjoint, voir par exemple [Chavent et al., 1975]. La mise en œuvre de cette méthode est cependant extrêmement délicate pour les problèmes complexes que sont les écoulements tri-phasiques compositionnels, c'est-à-dire lorsque les différentes phases (gaz, eau, huile) peuvent se mélanger.

Cependant, les méthodes d'optimisation ont leurs limites dans la mesure où elles ne fournissent qu'une unique réalisation calée dépendant de son point de départ. On peut bien sûr faire varier le point de départ et obtenir ainsi de multiples réalisations, ce qui est fait en pratique, mais il est impossible de déterminer en quelle mesure elles seront représentatives de la loi (1.3), puisqu'en outre, elles ne prennent pas en compte les éventuelles erreurs de mesure sur les données. En général, dans les problèmes d'intégration des données de production, la fonction objectif (1.6) peut avoir un grand nombre de minima isolés, voir [Watkins, 1993]. Autrement dit, la distribution de probabilité (1.3) peut être multi-modale. Il est alors nécessaire de trouver une méthode capable d'identifier ces différents minima et de leur associer un poids, autrement dit une certaine probabilité d'occurrence ; ceci dans la perspective d'évaluer les incertitudes portant sur les prévisions effectuées selon différentes réalisations issues de (1.3). Dans cette optique, une méthode consistant à réaliser un ensemble d'optimisations a été développée dans les années 90, voir [Kitanidis, 1995] et [Oliver, 1996]. L'idée est de minimiser (1.6) un certain nombre de fois en perturbant la donnée d relativement à l'erreur de mesure admise. Le nombre d'optimisations est alors choisi en fonction des capacités informatiques disponibles en considérant des points de départs différents. Il est néanmoins difficile dans ce cadre d'évaluer de quelle manière le résultat sera biaisé par le choix des points de départs de l'algorithme d'optimisation et le nombre total d'optimisations effectuées. En outre, cette méthode n'évacue pas les problèmes liés à l'emploi des méthodes d'optimisation déterministes notamment la présence éventuelle de minima locaux.

Pour éviter les problèmes de minima locaux, la méthode d'optimisation stochastique de recuit simulé a été implémentée plusieurs fois, voir [Deutsch and Journal, 1994] pour une revue. Pour autant, la lenteur de la convergence de cette méthode en regard des méthodes d'optimisation ne justifie pas son emploi.

Au vu des défauts des méthodes d'optimisation, notamment leur incapacité à évaluer les incertitudes liées au problème de calage d'historique, il semble plus pertinent de chercher à simuler directement des réalisations conditionnées aux données de productions, *i.e* suivant la loi (1.3). Les méthodes de Monte-Carlo par chaînes de Markov (McMC, voir partie 3 et plus généralement [Robert and Casella, 2004]) permettent en théorie de générer des échantillons d'une loi de probabilité sans limitation sur celle-ci. Elles sont donc particulièrement adaptées aux problèmes d'inférence bayésienne, notamment (1.3). Dans le domaine pétrolier, la première application de ces méthodes est due à [Tjelmeland et al., 1994]. En s'inspirant de ce travail, [Oliver et al., 1997] présente une application plus concrète de leur méthode. Plus récemment, l'algorithme de Langevin a été implémenté dans [Dostert et al., 2006] et une méthode adaptative dans [Holden et al., 2002] dans la problématique du calage d'historique. Pour autant, malgré leurs bonnes propriétés théoriques, ces méthodes présentent certains problèmes pratiques : du fait du coût informatique d'une simulation d'écoulement et de la dimension du problème, l'exploration de la loi *a posteriori* (1.3) par la chaîne est généralement incomplète en un temps raisonnable, il arrive souvent qu'elle se limite à la description d'un seul mode, qu'elle ne converge pas ou encore qu'elle reste bloquée dans un équilibre méta-stable, équivalent d'un minimum local dans le cas des algorithmes de simulation.

1.4 Objectifs et résultats

1.4.1 Objectifs

La problématique de la thèse est celle du calage d'historique des modèles géostatistiques et l'évaluation des incertitudes associées. Le principal objectif est de développer une méthodologie pour résoudre le problème d'inférence bayésienne (1.3) en prenant en compte les différents écueils évoqués dans la littérature et passés en revue dans le paragraphe 1.3.3.

1.4.2 Plan

Le présent manuscrit, outre cette introduction, est composé de trois chapitres. Dans les deux premiers, nous traitons deux aspects différents du problème de calage d'historique. Le premier est celui de la paramétrisation et de la réduction de la dimension du problème inhérente, le second est la problématique de l'échantillonnage ; ces deux chapitres peuvent être lus indépendamment. Le dernier chapitre présente des applications des méthodes abordées dans les deux premiers.

Dans le chapitre 2, nous présentons une méthode de réduction de la dimension basée sur la décomposition de Karhunen-Loève. Nous en décrivons les fondements théoriques et la mise en application à la paramétrisation des modèles géostatistiques, aussi bien pour les modèles continus que pour les modèles en faciès. Nous validons également

cette approche par une évaluation du nombre de composantes nécessaires à reproduire le comportement dynamique du champ sur un problème d'écoulement.

Le troisième chapitre est dévolu à la présentation des méthodes de Monte-Carlo par chaînes de Markov que nous employons pour résoudre le problème d'inférence bayésienne (1.3). Après quelques rappels théoriques, nous évaluons les principaux défauts de ces méthodes, en particulier pour la simulation de lois multi-modales. Nous introduisons alors le principe des chaînes de Markov en interaction et discutons de ses avantages sur les méthodes classiques.

Le quatrième chapitre consiste en une compilation de différents articles rédigés durant la thèse, qui présentent les résultats de la méthodologie développée, basée sur la réduction de la dimension par la décomposition de Karhunen-Loève et l'emploi des méthodes de chaînes de Markov en interaction.

1.4.3 Résultats et perspectives

Cette thèse est un travail de mathématiques appliquées. A ce titre, les résultats obtenus ne sont pas d'ordre théorique. Comme évoqué ci-dessus, l'objectif principal est de développer une méthodologie pour la résolution du problème de calage d'historique dans le cadre bayésien, c'est-à-dire un échantillonnage efficace de (1.3). Le fruit de ce travail prend ainsi la forme de résultats de simulations numériques, dont les performances sont comparées, le cas échéant, avec celles des méthodologies déjà mises en œuvre sur le sujet.

Le point de départ de cette thèse fut la méthode de déformation graduelle ([Hu, 2000]) abordée au paragraphe 2.6.1, page 78. En cherchant un critère pour déterminer le nombre de paramètres nécessaires pour échantillonner de manière satisfaisante la loi a posteriori (1.3), nous nous sommes orientés vers la décomposition de Karhunen-Loève du modèle géostatistique. Elle consiste à représenter celui-ci dans la base formée des vecteurs propres de son opérateur de covariance. Dans sa version continue, on obtient ainsi une représentation fonctionnelle du champ aléatoire. En ne conservant qu'un nombre fini de composantes dans cette représentation, on obtient une approximation du champ aléatoire qui représente une part quantifiable de la variance du processus. En effet, à chaque terme de la décomposition est attribuée une part de la variance globale qui est égale à la valeur propre associée au vecteur propre correspondant. En outre, l'approximation ainsi construite est optimale au sens de l'erreur quadratique moyenne intégrée (voir paragraphe 2.3.1 p. 36). On peut ainsi quantifier l'erreur d'approximation en terme de variance et construire éventuellement des zones de confiance autour des approximations.

On a étudié le comportement de ces approximations pour les modèles de covariance courants en géostatistique. Une étude a été également menée pour les modèles en faciès. Cette paramétrisation a été validée sur des problématiques d'écoulement à la fois dans le cas des modèles continus et des modèles en faciès. On a ainsi justifié l'emploi de cette méthode pour réduire la dimension du problème d'inversion lié à l'intégration des données de production. On a appliqué une approche d'analyse de sensibilité portant sur les composantes de cette décomposition sur un cas synthétique de réservoir, dont les résultats ont fait l'objet d'un article dont une version préliminaire est reproduite

page 160.

On a également développé un algorithme en vue d’une application pratique de cette décomposition sur un cas réel de réservoir. En effet, la décomposition de KL nécessite la diagonalisation de la matrice de covariance, la dimension de celle-ci étant le carré du nombre de points de la grille de discrétisation. Cette étape est donc impraticable directement pour des champs comprenant parfois plus de 10^6 points de discrétisation. Cet algorithme, présenté au paragraphe 2.6.2 page 80, passe par un découpage du champ en blocs dont la taille autorise l’application de la décomposition. Une analyse de ses propriétés de stabilité reste néanmoins à réaliser. Dans le cadre de cet algorithme, il serait intéressant d’étudier finement, au sein de chaque bloc, le nombre de composantes à employer pour reproduire les écoulements de manière satisfaisante, notamment en fonction de la position de ces blocs vis-à-vis des puits : l’écoulement devrait être plus sensible aux hétérogénéités de faible amplitude aux abords des puits. Il serait également intéressant de faire le lien entre cette décomposition et les méthodes de changement d’échelles, voir *e.g.* [Nøtinger et al., 2005].

La seconde partie du travail mené durant cette thèse a concerné l’échantillonnage de la loi conditionnelle (1.3), page 23. Une simulation d’écoulement étant fortement coûteuse en temps, il convient d’échantillonner de manière parcimonieuse tout en balayant au maximum l’espace d’incertitude, de très grande dimension malgré une éventuelle réduction, le but étant de construire un échantillon représentatif de la loi a posteriori. Cet objectif est éminemment contradictoire : il est clair que pour qu’un échantillon d’une loi de probabilité, sur un espace de grande dimension, soit représentatif, il est nécessaire de produire un grand nombre de points, ce qui se traduit par autant de simulations d’écoulement dans la cadre de notre problème. En outre, il est impossible de générer directement des points distribués selon (1.3), des simulations supplémentaires sont donc nécessaires. Nous nous sommes ainsi concentré sur le développement d’une méthode d’échantillonnage efficace visant à produire un échantillon distribué approximativement selon (1.3).

Les méthodes les plus efficaces pour échantillonner ce type de loi, malgré leur coût informatique, semblent être les méthodes de Monte Carlo par chaîne de Markov (McMC). Dans un premier temps, nous avons étudié différents algorithmes McMC, dont nous avons mis en évidence les carences, lorsqu’appliqués à l’échantillonnage de lois multimodales (voir section 3.5, p. 109). En effet, il existe 3 principaux écueils dans la pratique des algorithmes McMC sur les problèmes de grande dimension : la lenteur de l’exploration de l’espace d’état, le faible taux d’acceptation des transitions et la longueur des autocorrélations le long de la chaîne.

Pour pallier ces défauts, nous avons abordé des méthodes récemment décrites dans la littérature, consistant à faire interagir un certain nombre de chaînes de Markov lancées en parallèle. Les schémas d’interaction étudiés sont au nombre de trois : *parallel tempering*, *equi-energy sampler* et *population importance re-sampling*. Les résultats de chacune de ces méthodes sur un exemple simple ont été satisfaisants (paragraphe 3.6.6, p. 124).

On a alors appliqué la méthode de *population importance re-sampling*, associée à la réduction de dimension basée sur la décomposition de KL, à un cas synthétique de calage d’historique. Cet algorithme a été choisi en raison de propriétés supérieures aux

1.4. OBJECTIFS ET RÉSULTATS

deux autres, notamment une convergence plus rapide vers le régime stationnaire, et une réflexion a été menée quant au réglage des ses paramètres. Une application a été réalisée sur le problème du conditionnement d'un champ de perméabilité à des données de water cut (proportion d'eau dans l'huile produite) et a été soumis en vue d'une publication dans la revue *Computational Geosciences*, la dernière version de cet article est reproduite *in extenso* à la page 127. Une seconde application de notre méthodologie a été effectuée sur le conditionnement d'un modèle en faciès à des données de production ; l'on y a adjoint une comparaison avec les résultats obtenus par une méthode McMC classique. Ce travail a également fait l'objet d'une publication à paraître dans les actes du congrès ECMOR XI, Bergen, Norway, dont l'on trouvera copie en page 150.

Chapitre 2

Décomposition de Karhunen-Loève : théorie et application à l'ingénierie de réservoir

2.1 Introduction

Les modèles géostatistiques utilisés en ingénierie de réservoir pour modéliser les grandeurs de perméabilité et de porosité des roches, sont pour la plupart basés sur des fonctions aléatoires gaussiennes d'ordre 2, indexées sur l'espace physique du réservoir. Ils sont discrétisés sur un maillage généralement cartésien. Ils sont donc entièrement déterminés par leur espérance et leur fonction de covariance. On dressera la liste des propriétés intéressantes de ces modèles dans la section 2.2.

Dans le problème d'inversion bayésienne décrit en introduction, il s'agit de conditionner un champ aléatoire, discrétisé sur une grille, aux données de production observées. La dimension de ce problème est donc le nombre de points de la grille : la simulation d'un champ aléatoire sur une grille nécessite en effet l'emploi d'autant de variables aléatoires que de points de discrétisation, au nombre de 10^5 à 10^6 dans les applications pratiques. Il est ainsi souhaitable de réduire la dimension de ce problème par une paramétrisation adaptée. Nous présentons dans ce chapitre une méthode de réduction de la dimension basée sur la décomposition de Karhunen-Loève [Loève, 1955] du modèle géostatistique. Elle consiste à représenter le champ aléatoire dans la base formée des vecteurs propres de son opérateur de covariance. On obtient ainsi une représentation fonctionnelle du champ. Cette méthode a été évoquée dès les années 70 dans le domaine pétrolier (cf. [Gavalas et al., 1976]), pour autant un grand nombre de ses propriétés n'ont, à notre connaissance, pas été valorisées. En ne conservant qu'un nombre limité de composantes dans cette représentation, on obtient une approximation du champ aléatoire qui représente une part quantifiable de la variance du processus. En effet, à chaque terme de la décomposition est attribuée une part de la variance globale qui est égale à la valeur propre associée au vecteur propre correspondant. On peut ainsi quantifier l'erreur d'approximation en terme de variance et construire des zones de confiance autour des approximations. Cette méthodologie est décrite dans la section 2.3 du chapitre.

On a étudié le comportement de ces approximations pour les modèles de covariance courants en géostatistique décrivant des grandeurs continues (porosité, perméabilité)

aussi bien d'un point de vue statistique qu'au niveau du comportement dynamique. Cette étude fait l'objet de la section 2.4 du présent chapitre.

Une étude a été également menée pour les modèles dits en faciès, où les réalisations gaussiennes sont tronquées pour représenter la dispositions des différents types de roches présents dans le réservoir et auxquels sont ensuite allouées des valeurs de perméabilité et de porosité. Les résultats obtenus sont présentés dans la partie 2.5.

Enfin, nous présentons en partie 2.6 quelques applications de cette méthode.

2.2 Quelques rappels sur les processus du second ordre

Une grande partie de cette section est inspirée de l'ouvrage de Daniel Revuz ([Revuz, 1997]).

Dans un premier temps, il est essentiel de remarquer que les modèles géostatistiques de réservoir "basés sur le variogramme" sont des processus du second ordre, indexés généralement sur un compact \mathbf{U} inclus dans \mathbb{R}^3 . Dans la suite, \mathbf{U} désignera un espace discret, dénombrable ou continu de dimension quelconque. Dans un souci de simplification, on identifiera une variable aléatoire (v.a.) avec sa classe d'équivalence presque sûre.

Définition 2.2.1 *Un processus $X = (X_u)$, $u \in \mathbf{U}$, à valeurs réelles est dit du second ordre si toutes les variables X_u sont de carré intégrable.*

Il est dit gaussien si pour tout sous-ensemble fini (u_1, u_2, \dots, u_n) de \mathbf{U} la variable vectorielle $(X_{u_1}, X_{u_2}, \dots, X_{u_n})$ est gaussienne. Le processus est dit centré si $\mathbb{E}[X_u] = 0$ pour tout $u \in \mathbf{U}$.

On remarque qu'un processus gaussien est toujours du second ordre.

Définition 2.2.2 *Si X est un processus du second ordre, sa covariance est la fonction définie sur $\mathbf{U} \times \mathbf{U}$ par :*

$$\Gamma(u, v) = \text{Cov}(X_u, X_v)$$

On remarque qu'elle est toujours bien définie et finie. Cette fonction de covariance présente un certain nombre de propriétés remarquables — le lecteur intéressé pourra se rapporter à [Revuz, 1997] pour les démonstrations — parmi les plus intéressantes pour notre propos, citons :

1. La covariance d'un processus du second ordre est symétrique et semi-définie positive.
2. Réciproquement, toute fonction symétrique et semi-définie positive est la covariance d'un processus gaussien.
3. Pour tout processus du second ordre, il existe un processus gaussien de même covariance.
4. Si une mesure μ est symétrique sur \mathbb{R} et ν sa transformée de Fourier, la fonction définie sur \mathbb{R}^2 par :

$$\Gamma(u, v) = \nu(|v - u|)$$

est semi-définie positive. Le processus gaussien centré correspondant est dit *stationnaire*.

Définition 2.2.3 *Un processus X indexé par $u \in \mathbf{U}$ est dit stationnaire si quel que soit $n \in \mathbb{N}$, quels que soient $u_1, u_2, \dots, u_n, k \in \mathbf{U}$, $\mathbb{P}_{u_1, u_2, \dots, u_n} = \mathbb{P}_{u_1+k, u_2+k, \dots, u_n+k}$, autrement dit, si la distribution de ses marges finies dimensionnelles sont invariantes par translation.*

Définition 2.2.4 *Un processus X indexé par $u \in \mathbf{U}$ défini sur $(\Omega, \mathcal{A}, \mathbb{P})$ sera dit mesurable si l'application $(u, \omega) \rightarrow X_u(\omega)$ est mesurable par rapport à la tribu produit $\mathcal{B}(\mathbf{U}) \otimes \mathcal{A}$. Il sera dit continu si ses trajectoires sont p.s. continues, c'est-à-dire si pour presque tout ω , l'application $u \rightarrow X_u(\omega)$ est continue.*

Nous donnons maintenant les définitions de la continuité et de la différentiabilité en moyenne quadratique, qui sont bien adaptées aux processus du second ordre.

Définition 2.2.5 *Un processus X indexé par \mathbf{U} défini sur $(\Omega, \mathcal{A}, \mathbb{P})$ sera dit continu en moyenne quadratique si :*

$$\forall (u, v) \in \mathbf{U} \times \mathbf{U}, \lim_{v \rightarrow u} \mathbb{E} (X_u - X_v)^2 = 0$$

Il sera dit différentiable en moyenne quadratique si :

$$\forall (u, h) \in \mathbf{U} \times \mathbf{U}, \lim_{h \rightarrow 0} \mathbb{E} \left(\frac{X_u - X_{u+h}}{h} \right)^2 \text{ existe et est finie}$$

On peut remarquer que, si le processus est uniformément intégrable, c'est-à-dire si $\sup_U \mathbb{E}(|X_u|^{2+\varepsilon}) < \infty$, la continuité presque sûre implique la continuité en moyenne quadratique. En revanche, la réciproque est fautive. Le processus de Poisson homogène (N_u) d'intensité $\lambda \in \mathbb{R}$ en dimension 1 en est un bon contre-exemple : on sait que $\forall \omega \in \Omega$, les trajectoires de $N_u(\omega)$ sont discontinues, en revanche, on a :

$$\mathbb{E} (N_u - N_{u+h})^2 = \lambda h (1 + \lambda h)$$

et donc les trajectoires du processus de Poisson homogène en dimension 1 sont continues en moyenne quadratique.

Nous en arrivons maintenant à la proposition fondamentale qui va nous permettre de justifier la décomposition de Karhunen-Loève :

Proposition 2.2.1 *Soit X_u , $u \in \mathbf{U}$ un processus du second ordre centré, réel, de covariance Γ . Il existe un espace de Hilbert \mathcal{H} de fonctions réelles sur \mathbf{U} telles que :*

i) l'espace vectoriel \mathcal{H}_0 engendré par les fonctions $\Gamma(u, \cdot)$ $u \in \mathbf{U}$ est dense dans \mathcal{H} ,

ii) pour toute fonction h de \mathcal{H} ,

$$h(u) = \langle h, \Gamma(u, \cdot) \rangle$$

où $\langle h(\cdot), g(\cdot) \rangle = \int_{\mathbf{X}} h(x)g(x)dx$ désigne le produit scalaire usuel sur \mathcal{H} .

Cet espace s'appelle l'espace auto-reproduisant (EAR) de X ou de Γ . Il est isomorphe au sous-espace \mathbf{H} de $L^2(\Omega, \mathcal{A}, \mathbb{P})$ engendré par les (classes des) variables aléatoires X_u , $u \in \mathbf{U}$, dans l'isométrie qui prolonge l'application $\Gamma(u, \cdot) \rightarrow X_u$.

Le nom d'espace auto-reproduisant provient de la propriété :

$$\Gamma(u, v) = \langle \Gamma(u, \cdot), \Gamma(v, \cdot) \rangle .$$

2.2. QUELQUES RAPPELS SUR LES PROCESSUS DU SECOND ORDRE

On peut d'ailleurs remarquer que l'on peut énoncer la majeure partie du résultat précédent sans faire allusion aux processus et que c'est en fait principalement un résultat sur les fonctions semi-définies positives.

L'isomorphisme entre \mathbf{H} et \mathcal{H} présente en particulier l'intéressante propriété que l'image d'une base orthogonale par cet isomorphisme est une base orthogonale.

Proposition 2.2.2 *Si $(\xi_i), i \in I$, est une base orthonormale de \mathbf{H} et $(\phi_i), i \in I$, la base correspondante de \mathcal{H} , alors pour tout $u \in \mathbf{U}$,*

$$X_u(\omega) = \sum_{i \in I} \xi_i(\omega) \phi_i(u) \text{ et } \Gamma(u, \cdot) = \sum_{i \in I} \phi_i(u) \phi_i, \omega \in \Omega,$$

où les séries convergent respectivement dans \mathbf{H} et dans \mathcal{H} . De plus, pour tout couple (u, v) de $\mathbf{U} \times \mathbf{U}$,

$$\Gamma(u, v) = \sum_{i \in I} \phi_i(u) \phi_i(v).$$

On trouvera une démonstration de cette proposition dans le cas gaussien dans [Revuz, 1997]. Lorsque X est gaussien centré, l'espace \mathbf{H} est ce que l'on appelle un espace gaussien i.e. un sous-espace vectoriel de L^2 tel que toute sous-famille finie de v.a. de cet espace soit gaussienne vectorielle. Dans ce cas, les variables ξ_i du résultat précédent étant deux à deux orthogonales, sont en fait indépendantes. Ceci mène à une autre construction des processus gaussiens qui n'utilise que le théorème d'existence des familles de variables indépendantes.

Proposition 2.2.3 *Si Γ est une covariance sur $\mathbf{U} \times \mathbf{U}$ et si (ϕ_i) est une famille orthogonale dans \mathcal{H} , pour toute famille (ξ_i) de variables indépendantes et de loi $\mathcal{N}(0, 1)$, le processus X défini par*

$$X_u(\omega) = \sum_{i \in I} \xi_i(\omega) \phi_i(u), \omega \in \Omega,$$

est un processus gaussien de covariance Γ .

Il faut comprendre que, $(\Omega, \mathcal{A}, \mathbb{P})$ étant l'espace sur lequel sont définies les ξ_i , les familles $\xi_i \phi_i(u)$ sont sommables dans l'espace $L^2(\mathbb{P})$ et la sommation définit donc une classe de v.a. parmi laquelle on choisit une variable X_u .

2.3 Théorème de Karhunen-Loève

Un choix judicieux de la base (ϕ_i) et de X_u dans sa classe d'équivalence permet d'obtenir des versions ayant de bonnes propriétés. Notamment, le théorème de Karhunen-Loève (KL) [Loève, 1955] propose d'utiliser la famille des fonctions propres de $\Gamma(u, \cdot)$ comme famille orthonormale de \mathcal{H} . En effet, $\Gamma(u, \cdot)$ étant semi-définie positive, le théorème de Mercer [Mercer, 1909] sur les noyaux semi-définis positifs assure l'existence et l'unicité de cette décomposition.

2.3.1 Enoncé du théorème

Proposition 2.3.1 (Théorème de Karhunen-Loève) *Soit X_u , $u \in \mathbf{U}$ un processus du second ordre centré, réel, de covariance Γ . Si Γ est une covariance continue sur $\mathbf{U} \times \mathbf{U}$, il existe (e_i) une base orthonormale de \mathcal{H} , constituée des fonctions propres de $\Gamma(u, \cdot)$. La base correspondante de \mathbf{H} est la famille (Z_i) de variables non-corrélées, centrées. En particulier, on a les représentations suivantes :*

- i) $Z_i(\omega) = \int_{\mathbf{U}} X_u(\omega) e_i(u) du$,
- ii) $X_u(\omega) = \sum_{i \in I} Z_i(\omega) e_i(u)$, $\omega \in \Omega$,
- iii) $\text{Var}(Z_i) = \lambda_i$, où λ_i désigne la valeur propre associée à la fonction propre $e_i(u)$.

Dans le cas d'un processus gaussien, on a en outre :

Proposition 2.3.2 *Soit X_u , $u \in \mathbf{U}$ un processus gaussien centré, réel, de covariance Γ continue sur $\mathbf{U} \times \mathbf{U}$. Alors, dans le théorème précédent, les (Z_i) sont indépendantes gaussiennes, centrées de variance λ_i . Soit (ξ_i) une famille de variables indépendantes de loi $\mathcal{N}(0, 1)$, X_u admet la représentation suivante ,*

$$X_u(\omega) = \sum_{i \in I} \sqrt{\lambda_i} \xi_i(\omega) e_i(u), \quad \omega \in \Omega, \quad (2.1)$$

De plus si l'on identifie I à l'espace des entiers naturels non nuls \mathbb{N}^* , ce qui est justifié en toute généralité par le fait que I est au plus dénombrable, on a le résultat suivant :

$$\lim_{N \rightarrow \infty} \sum_{i=1}^N \sqrt{\lambda_i} \xi_i(\omega) e_i(u) = X_u(\omega), \quad \omega \in \Omega, \quad (2.2)$$

presque sûrement.

L'intérêt de la proposition précédente dans notre problème est d'employer une somme finie de (ξ_i) pour représenter X_u par un processus gaussien approché :

$$\hat{X}_u^N(\omega) = \sum_{i=1}^N \sqrt{\lambda_i} \xi_i(\omega) e_i(u), \quad \omega \in \Omega, \quad (2.3)$$

Cette formule nous donne une représentation approximative d'un processus gaussien, utilisant un nombre fini de v.a. gaussiennes. On peut d'ailleurs démontrer que cette représentation est optimale au sens de l'erreur quadratique moyenne ($\varepsilon_M^2 = \mathbb{E}(\hat{X}_u^N - X_u)^2$) intégrée sur le domaine \mathbf{U} [Ghanem and Spanos, 1991].

2.3. THÉORÈME DE KARHUNEN-LOÈVE

Enfin, on peut citer un résultat intéressant d'un point de vue purement statistique (cf. [Pycke, 2003]). Sous les hypothèses de la proposition 2.3.2, on a le résultat suivant :

Lemme 2.3.1 *Soit X_u , $u \in \mathbf{U}$ un processus gaussien centré, réel, de covariance Γ continue sur $\mathbf{U} \times \mathbf{U}$. Si λ_1 désigne la plus grande valeur propre de son développement de Karhunen-Loève, on a lorsque x tend vers l'infini :*

$$\log(P(\int_{\mathbf{U}} X_u^2 du \geq x^2)) \approx -\frac{x^2}{2\lambda_1}$$

Ce résultat de *grande déviation* permet de contrôler l'erreur globale sur \mathbf{U} commise par l'approximation (2.3) :

$$\log(P(\int_{\mathbf{U}} (\hat{X}_u^N - X_u)^2 du \geq x^2)) \approx -\frac{x^2}{2\lambda_{N+1}}$$

Une autre propriété intéressante de cette représentation pour notre problème est la facilité de changer l'échelle de la grille sur laquelle est discrétisé le modèle, une fois les fonctions de base connues. On a alors une représentation fonctionnelle du processus gaussien, ce qui nous permet pour passer d'une grille fine à une grille plus grossière et inversement, en considérant la valeur de ces fonctions aux noeuds de la grille de discrétisation considérée.

Dans la suite, dans un souci de simplification, on ne reproduira plus l'aléa $\omega \in \Omega$ dans les différentes formules.

2.3.2 Conservation de la variance

Pour autant, la représentation (2.3) ne conserve pas la variance des marginales du processus. En effet, on a :

$$\text{Var}(\hat{X}_u^N) = \text{Var}(\sum_{i=1}^N \sqrt{\lambda_i} \xi_i e_i(u)) = \sum_{i=1}^N \lambda_i \neq \text{Var}(X_u) = \sum_{i=1}^{\infty} \lambda_i$$

On peut pallier ce problème, en prenant :

$$\tilde{X}_u^N = \sqrt{\frac{\text{Var}(X_u)}{\sum_{i=1}^N \lambda_i}} \sum_{i=1}^N \sqrt{\lambda_i} \xi_i e_i(u) \quad (2.4)$$

Cette représentation est clairement cohérente avec le résultat asymptotique de la proposition 5 et nous permet d'avoir la propriété suivante, pour tout u de \mathbf{U} et pour tout $N \in \mathbb{N}^*$:

$$\text{Var}(\tilde{X}_u^N) = \text{Var}(X_u)$$

C'est-à-dire que ce représentant de X_u conserve les moments d'ordre 2 des marginales unidimensionnelles du processus d'origine. En notant E_M^2 l'erreur quadratique moyenne de notre nouvelle représentation et en notant :

$$\sqrt{\frac{\text{Var}(X_u)}{\sum_{i=1}^N \lambda_i}} = 1 + a,$$

on a :

$$E_M^2 = \mathbb{E} \left[\tilde{X}_u^N - X_u \right] = a^2 \text{Var}(X_u) + \varepsilon_M^2 \quad (2.5)$$

On perd donc de cette manière l'optimalité au sens de l'erreur quadratique moyenne. On peut cependant remarquer que a^2 et ε_M^2 ont une limite égale à 0 lorsque N tend vers l'infini et sont du même ordre de grandeur ; on choisira l'une ou l'autre représentation à sa guise tout en gardant à l'esprit que la conservation de la variance présente un avantage non négligeable dans certaines applications. Pour autant, on gardera (2.3) comme représentation de référence dans la suite de l'exposé.

2.3.3 Le problème aux valeurs propres

Maintenant que nous avons posé les bases théoriques permettant de définir la décomposition de KL, nous allons nous attacher à décrire les problèmes pratiques de son utilisation. Le problème essentiel induit par cette représentation est le calcul des valeurs propres et fonctions propres associées. Il s'écrit sous la forme d'une équation intégrale homogène de Fredholm [Ghanem and Spanos, 1991] du second ordre :

$$\int_{\mathbf{U}} \Gamma(u, v) e(u) du = \lambda e(v) \quad (2.6)$$

Ce type d'équation a été largement étudié et il existe une importante littérature à son sujet, dans laquelle on peut trouver les propriétés qui s'y rattachent. En outre, Γ étant une fonction de covariance bornée, symétrique et semi-définie positive, à support compact dans notre champ d'application, cela simplifie l'analyse du problème. En effet, il s'ensuit un certain nombre de propriétés pour les fonctions propres et les valeurs propres, parmi lesquelles on peut citer :

1. La famille $(e_i(u))$ de fonctions propres forme une base dénombrable orthonormale complète de \mathcal{H} ,
2. A chaque valeur propre λ_i correspond au plus un nombre dénombrable de fonctions propres linéairement indépendantes.
3. Pour tout $i \in \mathbb{N}^*$, $\lambda_i \in \mathbb{R}^+$.

On peut noter que ces propriétés sont indépendantes de la nature du processus sous-jacent ; la décomposition de KL peut ainsi être appliquée, en théorie, à une large gamme de processus du second ordre : non-stationnaires, non-gaussiens et/ou multidimensionnels. Attachons-nous maintenant à décrire les méthodes de résolution de (2). On peut dans certains cas trouver une solution analytique à ce problème mais dans la pratique, on aura le plus souvent recours à des schémas numériques.

Solution analytique de l'équation intégrale de Fredholm Pour certaines classes de processus, on peut trouver une solution analytique à l'équation (2.6). La méthode consiste à différentier (2.6) deux fois par rapport à v . On réinjecte l'expression trouvée dans (2.6) et on peut déterminer les valeurs propres en satisfaisant aux conditions aux limites données par les bornes de l'espace d'indexation du processus. Cette méthode se

2.3. THÉORÈME DE KARHUNEN-LOÈVE

limite cependant aux processus indexés sur un espace de dimension 1. Parmi les classes de processus pour lesquelles cette solution est calculable, citons :

- le processus de Markov du premier ordre en dimension 1 sur $[0,1]$, dont la fonction de covariance est l'exponentielle :

$$\Gamma(u, v) = e^{-|u-v|/a},$$

On définit ω_n , respectivement ω_n^* , comme étant la solution de la première, respectivement de la seconde équation du système :

$$\begin{cases} \frac{1}{a} - \omega \tan(\omega) = 0 \\ \omega + \frac{1}{a} \tan(\omega) = 0 \end{cases}$$

Les fonctions propres prennent alors la forme suivante :

$$\begin{aligned} e_i(u) &= \frac{\cos(\omega_n u)}{\sqrt{1 + \frac{\sin(2\omega_n)}{2\omega_n}}} \\ \text{et} \\ e_i^*(u) &= \frac{\cos(\omega_n^* u)}{\sqrt{1 - \frac{\sin(2\omega_n^*)}{2\omega_n^*}}}. \end{aligned}$$

pour n pair et impair respectivement. Les valeurs propres correspondantes sont :

$$\begin{aligned} \lambda_n &= \frac{2/a}{\omega_n^2 + 1/a^2} \\ \text{et} \\ \lambda_n^* &= \frac{2/a}{\omega_n^{*2} + 1/a^2} \end{aligned}$$

- le processus de Wiener ou mouvement Brownien (W_u) sur $[0, 1]$:

$$\Gamma(u, v) = \min(u, v),$$

dont le développement de KL s'écrit :

$$W_u = \frac{\sqrt{2}}{\pi} \sum_{i=1}^{\infty} \xi_i \left(\frac{2}{2i-1} \right) \sin \left(\frac{1}{2}(2i-1)\pi u \right), \quad u \in [0, 1]$$

- le pont Brownien (B_u) sur $[0,1]$:

$$\Gamma(u, v) = \min(u, v) - ts,$$

dont le développement de KL s'écrit :

$$B_u = \sum_{i=1}^{\infty} \frac{1}{i\pi} \xi_i \left\{ \sqrt{2} \sin(i\pi u) \right\}, \quad u \in [0, 1]$$

A noter que ces deux derniers sont des processus du second ordre non-stationnaires.

Schémas numériques de résolution de l'équation intégrale de Fredholm On peut citer différentes méthodes ou stratégies numériques pour résoudre l'équation (2.6) : la méthode de Nystrom [Press et al., 1992], les méthodes de type Galerkin ou Rayleigh-Ritz [Huang et al., 2001], [Phoon et al., 2004], ou encore [Chen et al., 2002a].

La première consiste simplement en le choix d'un schéma d'intégration numérique de (2.6) permettant de se ramener à un problème linéaire de calcul de valeurs propres et vecteurs propres. Dans [Press et al., 1992], les auteurs préconisent l'emploi d'un schéma de quadrature gaussienne de type Gauss-Legendre, eu égard au nombre de calculs impliqués. Cependant, les auteurs remarquent que cette méthode est limitée aux fonctions de covariance régulière, c'est-à-dire qui ne présentent pas de singularité particulière. Par exemple, elle n'est pas bien adaptée à la fonction de covariance exponentielle qui n'est pas dérivable en 0. Le problème aux valeurs propres est ensuite résolu via des méthodes de décomposition de matrices symétriques définies positives.

La seconde se présentent comme une famille de méthodes basées sur un principe commun qui peut se présenter en 3 étapes :

1. On choisit un ensemble fini (de taille K) de fonctions de base, notons les $\phi_1(u), \phi_2(u), \dots, \phi_K(u)$
2. On approxime chaque fonction propre par son projeté sur cette base :

$$e_i(u) = \sum_{k=1}^K d_{i,k} \phi_k(u), \quad (2.7)$$

3. On réécrit le système dans la nouvelle base comme suit : on injecte (2.7) dans (2.6) :

$$\sum_{k=1}^K d_{i,k} \int_{\mathbf{U}} \Gamma(u, v) \phi_k(u) du = \lambda_i \sum_{k=1}^K d_{i,k} \phi_k(v), \quad (2.8)$$

et pour $j = 1, \dots, K$, on projette (2.8) sur ϕ_j :

$$\sum_{k=1}^K d_{i,k} \int_{\mathbf{U}} \int_{\mathbf{U}} \Gamma(u, v) \phi_k(u) \phi_j(v) du dv = \lambda_i \sum_{k=1}^K d_{i,k} \int_{\mathbf{U}} \int_{\mathbf{U}} \phi_k(v) \phi_j(v) dv, \quad (2.9)$$

Et l'on peut réécrire cette 2.9 sous la forme d'un système linéaire :

$$AD = \Lambda BD, \quad (2.10)$$

où les composantes des matrices $K \times K$ sont :

$$A_{k,j} = \int_{\mathbf{U}} \int_{\mathbf{U}} \Gamma(u, v) \phi_k(u) \phi_j(v) du dv, \quad (2.11)$$

$$D_{i,k} = d_{i,k}, \quad (2.12)$$

$$\Lambda_{i,j} = \delta_{ij} \lambda_i, \quad (2.13)$$

$$B_{k,j} = \int_{\mathbf{U}} \int_{\mathbf{U}} \phi_k(v) \phi_j(v) dv, \text{ où } \delta_{ij} \text{ désigne le symbole de Kronecker.} \quad (2.14)$$

2.3. THÉORÈME DE KARHUNEN-LOÈVE

On résout ensuite le problème linéaire aux valeurs propres (2.10) pour déterminer les coefficients $d_{i,k}$ et les valeurs propres λ_i .

D'un point de vue pratique, on est amené à faire deux choix arbitraires : celui de la base de fonctions $\phi_k(x)$ et de l'entier K . Il est intéressant de choisir une base de fonctions orthonormale ; la matrice B est ainsi l'identité (cf. l'expression de ses coefficients) et le problème aux valeurs propres s'en trouve simplifié :

$$AD = \Lambda D.$$

On peut par exemple choisir une base de polynômes orthonormaux, une base d'ondelettes, ou encore une base d'ondes planes de Fourier. Le choix de la base tronquée influera directement sur la représentation obtenue, il pourra être intéressant de choisir une base adaptée au modèle de covariance considéré.

En ce qui concerne le choix de K , on trouve peu de renseignements dans la littérature si ce n'est que l'on préconise de prendre $K \gg N$ afin d'avoir une bonne représentation des fonctions propres. Le principal intérêt de cette méthode est de pouvoir contrôler l'erreur commise dans l'approximation de (2.6).

On peut voir en fait cette méthode comme une double approximation dont les propriétés de convergence sont démontrées aisément, voir [Ruiz-Molina et al., 1999]. Les auteurs présentent d'ailleurs dans cet article une version modifiée de la décomposition de KL permettant la différentiation des fonctions de base.

On peut encore citer d'autres méthodes : celle proposée dans [Schneider and Willsky., 2003], où les auteurs proposent un algorithme original permettant de calculer la décomposition de la covariance sur une base de sous-espaces de Krylov ; enfin on peut encore citer la méthode par collocation rapide proposée dans [Chen et al., 2002b].

Dans notre domaine d'étude, il convient de s'interroger sur la méthode à employer. En particulier, dans la mesure où les modèles de fonctions aléatoires d'ordre 2 sont discrétisés sur une grille, on leur associe une matrice de covariance de taille finie, qui correspond à la version discrétisée de la fonction de covariance. Est-il alors nécessaire de calculer une bonne approximation numérique des fonctions propres de l'opérateur continu de covariance ou bien suffit-il de diagonaliser la matrice de covariance et d'employer les vecteurs propres associés ? Nous répondrons à cette question dans la prochaine section, au moment d'aborder les applications sur les modèles usuels.

2.3.4 Relation avec la représentation harmonique

Une autre méthode de représentation de processus du second ordre, très populaire, est la méthode de représentation harmonique ou spectrale, c'est-à-dire la représentation de la covariance dans l'espace des ondes planes de Fourier. On emploiera ici le terme "harmonique" plutôt que "spectrale" pour éviter toute confusion avec la suite des valeurs propres de la fonction de covariance, appelée également spectre. Considérons, à titre d'exemple, un processus stationnaire (i.e $\Gamma(u, v) = \Gamma(|u - v|)$), du second ordre, indexé sur \mathbb{R} tout entier. Dans ce cas, l'équation (2.6) devient :

$$\int_{-\infty}^{+\infty} \Gamma(|v - u|) e(u) du = \lambda e(v) , \quad -\infty < v < +\infty \quad (2.15)$$

Ce problème se résout aisément en prenant $e(v) = e^{i\omega v}$. En substituant cette expression dans (2.15), on obtient :

$$\lambda = \int_{-\infty}^{+\infty} \Gamma(|v - u|) e^{-i\omega(v-u)} du = S(\omega). \quad (2.16)$$

Ainsi, les valeurs propres sont données par le spectre de puissance $S(\omega)$ où ω est la fréquence en radians. Dans le cas d'un processus sur un "long" intervalle $[-H, H]$, on peut faire l'approximation suivante :

$$\lambda_j \approx S(\omega_j) = S\left(\frac{\pi j}{H}\right), \quad (2.17)$$

$$e_j(u) \approx \frac{1}{\sqrt{2H}} e^{i(\pi j/H)u} \quad (2.18)$$

et l'on peut réinjecter ces expressions dans la représentation (2.3). De la sorte, on montre que, dans le cas d'un processus stationnaire du second ordre de longue durée, la décomposition de KL peut être approximée par la représentation harmonique. On se rapportera à [Huang et al., 2001], pour une étude numérique comparative dans le cas unidimensionnel avec une covariance de type exponentiel. Ils montrent principalement la supériorité de la représentation tronquée de la décomposition de KL par rapport à une représentation harmonique tronquée au même nombre de composantes.

La représentation harmonique est avantageuse dans le sens où l'on dispose d'algorithmes très rapides et efficaces pour la mettre en œuvre. Cependant, les performances de la représentation harmonique tronquée sont faibles par rapport à la décomposition de KL. Notre but étant de réduire le nombre de paramètres du modèle ; nous préférons la seconde. A noter que ce commentaire ne remet aucunement en cause la validité de la méthode FFT-MA [Le Ravalec et al., 2000]) employée à l'IFP pour simuler les champs aléatoires gaussiens. En effet, dans cette méthode, la base de fonctions harmoniques employée n'est pas tronquée, du moins pas dans le sens d'une approximation, et l'on obtient une représentation exacte du processus, respectivement au pas de discrétisation de la grille.

2.4 Application de la décomposition de Karhunen-Loève aux modèles de covariance courants en géostatistiques

Dans cette section, nous allons examiner les performances de la décomposition de KL comme méthode d'approximation des modèles géostatistiques. En particulier, nous allons chercher à déterminer en quelle mesure cette méthode permet de réduire la dimension du problème d'inversion. Dans un premier temps, nous allons nous attacher à décrire les modèles usuels en géostatistique pétrolière, analyser les résultats de la décomposition de ceux-ci en fonction de la portée et de la dimension du modèle et décrire un critère de choix du nombre de composantes de type statistique, typiquement le pourcentage de reproduction de la variance. Dans un second temps, nous allons confronter ce critère à des résultats dynamiques issus de simulations d'écoulement à différents ordres de troncature.

2.4.1 Analyse statistique

Présentation des modèles

Nous allons maintenant explorer plus en détail les intérêts de l'emploi de cette méthode dans le cadre des modèles géostatistiques continus. On s'attachera à déterminer, dans cette partie, pour différents modèles de covariance donnés, le nombre $N_{1-\alpha}$ de fonctions propres nécessaires à la description d'une part $(1 - \alpha)$ de la variance totale du processus. En particulier, on regardera l'influence de la dimension de l'espace d'indexation du processus, du pas de discrétisation employé pour les calculs des fonctions propres et la simulation, et du choix du nombre α , sur ce nombre $N_{1-\alpha}$. En ce qui concerne α , en adoptant la terminologie statistique, on le définira comme la marge d'incertitude ou bien $(1 - \alpha)$ désignera le niveau de confiance. On va se donner trois modèles de covariance classiques en géostatistiques pour effectuer ces tests numériques, à savoir :

1. le modèle exponentiel : $\Gamma(u, v) = \sigma^2 e^{-3 \frac{\|u-v\|}{a}}$,
2. le modèle normal ou gaussien : $\Gamma(u, v) = \sigma^2 e^{-3 \frac{\|u-v\|^2}{a^2}}$, on préférera le terme "normal" pour éviter la confusion avec la loi du processus,
3. le modèle sphérique : $\Gamma(u, v) = \sigma^2 (1 - 3 \frac{\|u-v\|}{2a} + \frac{\|u-v\|^3}{2a^3}) \mathbb{1}_{\{\|u-v\| < a\}}$, où $\mathbb{1}$ désigne la fonction indicatrice.

En géostatistique, on définit la portée pratique (pp) ou encore la longueur de corrélation comme la distance minimale au-delà de laquelle 2 points du processus seront "non-corrélés". Plus précisément, la définition de la portée n'a pas le même sens dans le cas du modèle sphérique où elle correspond à la distance h au-delà de laquelle la covariance s'annule, et le cas des modèles normaux et exponentiels, où la covariance à une limite égale à 0 lorsque h tend vers l'infini, c'est pourquoi l'on a besoin pour ces modèles de définir la portée pratique. Elle s'exprime en fonction du terme a dans les expressions précédentes. Dans le cas de la covariance sphérique, elle est exactement égale au terme a . Dans les cas de l'exponentielle et de la normale, la longueur de corrélation est définie

comme la distance minimale $\|u - v\|$ telle que $\Gamma(u, v) = \Gamma(\|u - v\|) \leq 0.05\Gamma(0)$, elle correspond en pratique au terme a dans l'expression de la covariance pour le modèle exponentiel ainsi que pour le modèle normal. Les expressions ci-dessus ont été construites de manière à obtenir des longueurs de corrélation équivalentes entre les 3 modèles.

En géostatistique, on considère généralement le variogramme plutôt que la covariance. En notant : $\Gamma(u, v) = \Gamma(\|u - v\|) = \Gamma(h)$, il se définit comme suit :

$$\gamma(h) = \Gamma(0) - \Gamma(h)$$

On représente sur la figure (2.1) les 3 variogrammes qui correspondent aux modèles considérés.

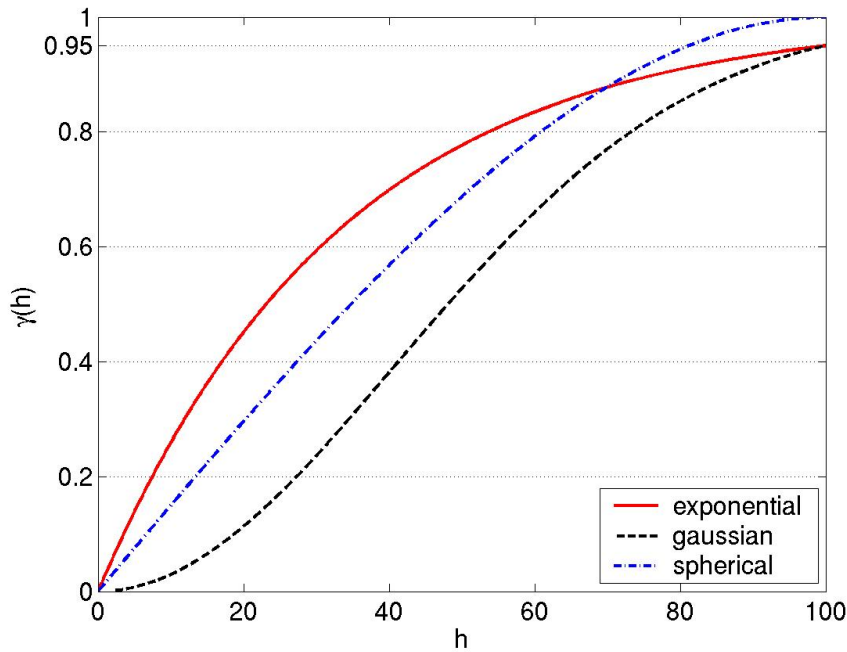


FIG. 2.1 – Variogrammes considérés.

On va également présenter les résultats sur les modèles de covariance précédents agrémentés d'un effet "pépité", c'est-à-dire d'une discontinuité à l'origine de la fonction de covariance. Dans ce cas, en notant p l'effet pépité, les modèles précédents se réécrivent :

$$\Gamma_{pep}(u, v) = \Gamma_{pep}(\|u - v\|) = p\mathbf{1}_{\{u=v\}} + \Gamma(u, v)$$

En vérité, on peut voir que ce modèle de covariance correspond à celui de la somme d'un processus muni d'un modèle de covariance continu et de variance $\Gamma(0)$ et d'un bruit blanc de variance p . Pour le voir, il suffit de poser $Y_u = X_u + \varepsilon_u$, avec X processus stationnaire du second ordre de covariance Γ et ε bruit blanc de variance p ; et de calculer la covariance de Y . On peut donc voir ces modèles comme des versions perturbées des précédents, où l'hétérogénéité des réalisations sera plus importante. A noter

que cette fonction de covariance n'est pas continue et sort donc du cadre d'application du théorème de Karhunen-Loève (cf. 2.3.1 et 2.3.2, p.36).

La méthode numérique employée ici est la méthode de Nystom décrite en 2.3.3 (p. 38) avec une discrétisation uniforme sur l'ensemble du domaine. On considère la covariance d'un processus 1D indexé sur $[0, H]$, celle d'un processus 2D sur $[0, H]^2$, et celle d'un processus 3D indexé sur $[0, H]^3$, $H = a$ ou $H = 4a$, où a désigne la portée pratique. Le maillage discrétisant est régulier et l'on fait varier son pas de $H/100$ à $H/3600$ pour le cas 1D, de $H/10$ à $H/60$ dans le cas 2D, sur l'espace d'indexation. On obtient ainsi des matrices de covariance dont les tailles sont égales pour les 2 cas, à chaque itération de notre algorithme. Dans le cas 3D, les calculs n'ont été effectués que pour les pas de discrétisation $H/10$ et $H/20$; au-delà, la matrice de covariance dépasse les capacités mémoire de *matlab*. En pratique, on pourra effectuer cette opération pour des champs bien plus grands moyennant l'implémentation de la méthodologie dans le langage adéquat, par exemple le *C*. Ensuite, le problème aux valeurs propres est résolu avec la fonction *matlab eig*, qui opère via une décomposition de Cholesky de la matrice considérée.

Résultats

On présente sur les figures suivantes (2.2 – 2.7) les sommes cumulées des valeurs propres, normalisées, *i.e.* divisées par la somme totale des valeurs propres, en fonction du logarithme du nombre de valeurs propres N dans les différents cas proposés. Dans les graphes suivants, les courbes sont tronquées à partir du moment où elles atteignent 99% de la variance totale ou bien à 1000 points si ce seuil n'est pas atteint dans les cas en 2 dimensions, et à 10000 points dans les cas en 3 dimensions. Dans le cas des modèles à effet pépité, on a représenté jusqu'à 10000 points de la courbe des sommes cumulées des valeurs propres en tronquant à 99% de la variance totale pour des raisons explicitées dans le commentaire. Ces derniers graphes se trouvent page 48.

La première chose qui nous apparaît au vu des différentes figures est l'impact du pas de discrétisation sur la valeur des valeurs propres calculées, particulièrement dans les cas 2D et 3D (2.4, 2.5, 2.6, 2.7). Pour chacun des modèles considérés, les courbes des sommes cumulées des valeurs propres s'infléchissent de plus en plus lorsque l'on diminue le pas de discrétisation, jusqu'à se stabiliser en deçà d'une certaine valeur du pas. Ce phénomène est d'une certaine manière mécanique : lorsque le pas diminue, la taille de la matrice de covariance et donc son nombre total de valeur propres augmentent. Quant à la stabilisation, elle provient simplement de la continuité de la covariance à l'origine. En effet, considérons un processus du second ordre X_u indexé sur \mathbf{U} compact de \mathbb{R}^d de fonction de covariance continue et stationnaire, *i.e.* $\Gamma(u, v) = \Gamma(\|u - v\|) = \Gamma(h)$, où $\|\cdot\|$ désigne la norme euclidienne. On a dans les cas considérés et, plus généralement, quel que soit le modèle de covariance continu à l'origine :

$$\mathbb{E}[(X_{u+h} - X_u)^2] = 2\Gamma(0) - 2\Gamma(h) = 2\gamma(h),$$

où $\gamma(h)$, le variogramme, est continu et tend vers 0 lorsque h tend vers 0, ce qui implique une limite nulle en 0 pour la quantité précédente. Sa décroissance est d'ailleurs

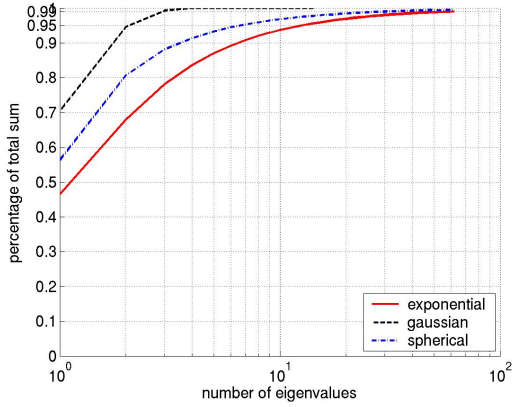


FIG. 2.2 – Somme cumulée des valeurs propres, cas 1D, $pp = H$.

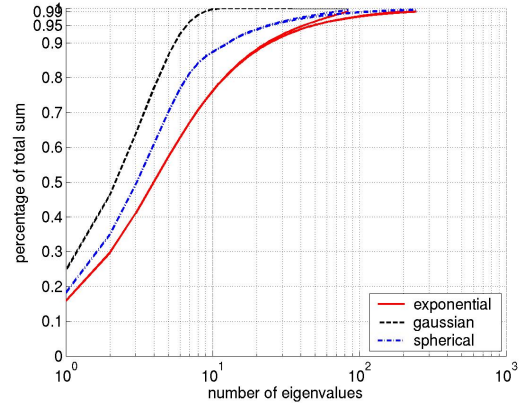


FIG. 2.3 – Somme cumulée des valeurs propres, cas 1D, $pp = H/4$.

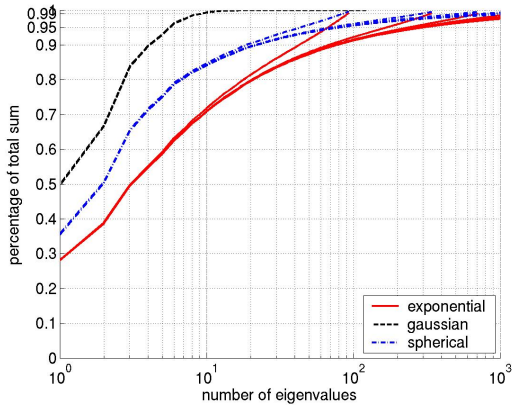


FIG. 2.4 – Somme cumulée des valeurs propres, cas 2D, $pp = H$.

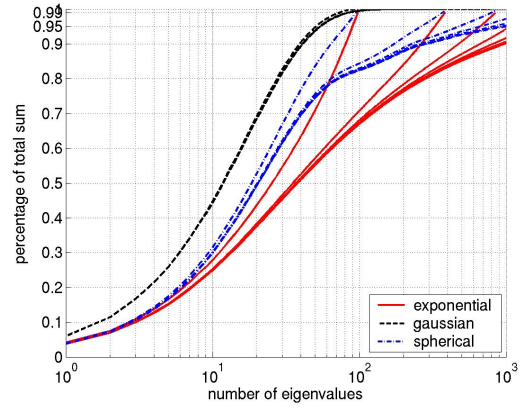


FIG. 2.5 – Somme cumulée des valeurs propres, cas 2D, $pp = H/4$.

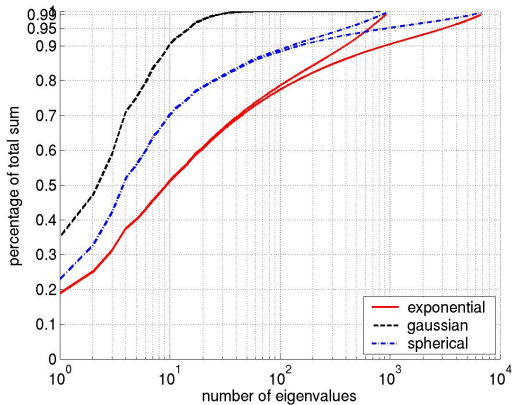


FIG. 2.6 – Somme cumulée des valeurs propres, cas 3D, $pp = H$.

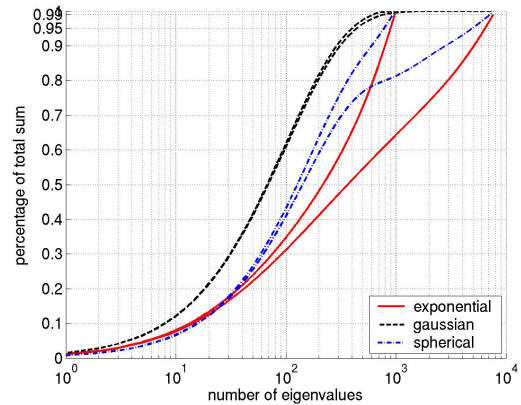


FIG. 2.7 – Somme cumulée des valeurs propres, cas 3D, $pp = H/4$.

d'autant plus rapide que la fonction de covariance est régulière. Ainsi, dans le cas de la fonction de covariance normale en 2 dimensions, un pas de $H/20$ semble suffisant

pour assurer la stabilité des N premières valeurs propres affichées, comme en attestent les graphiques. En effet, on a dans ce cas $\gamma(h) = 1 - e^{-h^2}$. Dans le cas des deux autres modèles, cette stabilisation est plus lente. Elle traduit le comportement de l'erreur due à la discrétisation de 2.6, p. 38. En vérité, on rejoint ici la problématique de la question posée en fin de la section 2.3.3. Au vu des résultats, dans l'optique de minimiser le nombre de paramètres nécessaires, il semble raisonnable de privilégier la seconde approche, c'est-à-dire de considérer les vecteurs propres correspondants à la matrice de covariance issue du maillage utilisé.

Dans la suite de la discussion, pour clarifier notre propos, on ne considérera que les sommes cumulées "stabilisées", c'est-à-dire au-dessous de ce pallier de discrétisation, pour les cas 1D et 2D et le tracé correspondant au pas de discrétisation $H/20$ dans le cas 3D.

On va observer précisément le comportement de $N_{1-\alpha}$ au niveau de confiance de 90%, i.e. $1 - \alpha = 0,9$. On peut voir sur les différentes figures (2.2 – 2.7) que $N_{1-\alpha}$ dépend avant tout du modèle de covariance considéré. Quelle que soit la dimension, le pourcentage de la somme cumulée des valeurs propres du modèle de covariance normal converge plus rapidement vers 1 que celle du modèle sphérique, qui converge plus vite que celle du modèle exponentiel. En ce qui concerne l'effet du passage de 1D à 2D puis à 3D, dans le cas $pp = H$ (figures 2.2, 2.4 et 2.6), on peut voir que lorsque l'on augmente la dimension du processus le nombre $N_{90\%}$ augmente fortement. Ainsi, pour le modèle normal, $N_{90\%}$ passe de 2 à 5 puis à 10 ; pour l'exponentiel, il passe de 7 à 78 puis à environ 900 ; enfin, pour le modèle de covariance sphérique, $N_{90\%}$ passe de 4 à 23 puis à environ 150.

Il est également remarquable que la convergence de la somme cumulée des valeurs propres est très lente, passé un certain seuil. Ainsi si l'on augmente le niveau de confiance de 90% à 95%, le nombre de valeurs propres à prendre en compte augmente très fortement, selon les modèles. En effet, pour le modèle normal, $N_{1-\alpha}$ passe de 2 à 3 en 1D, de 5 à 6 en 2D, et de 10 à 27 en 3D ; pour le modèle exponentiel, il passe de 7 à 14 en 1D, de 78 à environ 300 en 2D et approximativement de 900 à 3000 en 3D ; enfin pour le modèle sphérique, il passe de 4 à 7 en 1D, de 23 à 77 en 2D et de environ 150 à 800 en 3D.

Enfin, en ce qui concerne l'impact de la longueur du processus sur $N_{1-\alpha}$, la valeur augmente naturellement en fonction du ratio H/a de la longueur du processus (H) sur la longueur de corrélation (a). Cependant, l'augmentation du nombre de valeurs requises au niveau $1 - \alpha$ est beaucoup plus importante que celle du ratio H/a , ce phénomène est d'autant plus important quand la dimension augmente et varie selon le modèle considéré. Ainsi pour une multiplication du ratio par 4, en 1D (figures 2.2 et 2.3) pour le modèle normal, $N_{90\%}$ passe de 2 à 7 ; pour l'exponentiel, il passe de 3 à 9 et pour le sphérique de 4 à 11. En 2D (figures 2.4 et 2.5), pour le modèle normal, il passe de 5 à 50 ; pour le modèle exponentiel, il passe de 11 à 137 ; et pour le modèle sphérique, $N_{90\%}$ croît de 23 à 282. Enfin, pour les modèles en 3D (figures 2.6 et 2.7), on a une augmentation de 27 à environ 300 pour la covariance normale, de 900 à 5000 pour l'exponentielle et de 150 à 2000 pour la sphérique.

Observons maintenant l'impact d'un effet pépite ($p = 0.1$) sur $N_{90\%}$, pour les trois modèles de covariance étudiés, sur la longueur de corrélation : à première vue, les

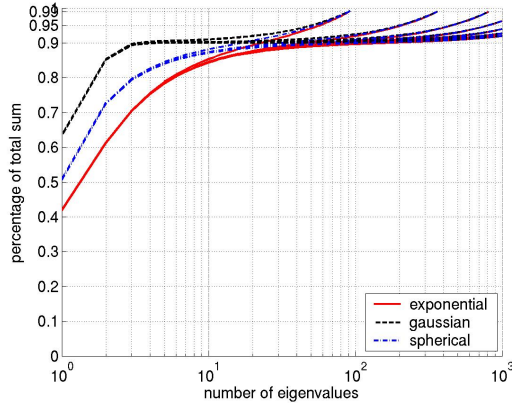


FIG. 2.8 – Somme cumulée des valeurs propres, cas 1D, effet pépité $p=0.1$, $pp = H$.

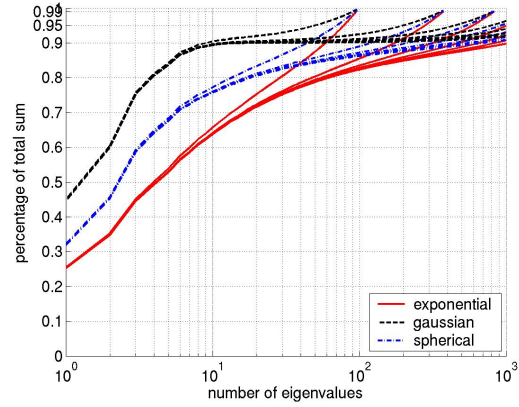


FIG. 2.9 – Somme cumulée des valeurs propres, cas 2D, effet pépité $p=0.1$, $pp = H$.

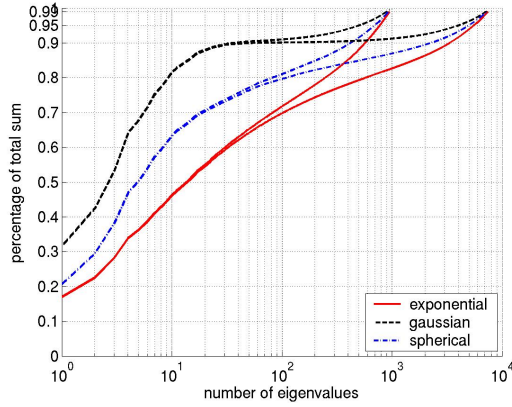


FIG. 2.10 – Somme cumulée des valeurs propres, cas 3D, effet pépité $p=0.1$, $pp = H$.

courbes de sommes cumulées des valeurs propres ne se stabilisent pas bien en fonction du pas de discrétisation (2.8,2.9,2.10), dans le sens vu plus haut, en présence d'un effet de pépité. En effet, comme la fonction de covariance présente une discontinuité en 0, on ne se trouve plus dans le cadre d'application du théorème de KL 2.3.1, page 36 . D'autre part, on peut voir que l'on a une convergence relativement rapide jusqu'au seuil de 90% de la variance, ce qui correspond à la part de la variance de la partie continue de la covariance (Z_t , de variance $\Gamma(0)$). Au-delà, la convergence est très lente et dans tous les cas de figure, il faut un nombre considérable de valeurs propres pour représenter 90% de la variance totale. Ce phénomène est simple à expliquer : si l'on considère un bruit blanc gaussien indexé sur T , la matrice de covariance des variables aléatoires qui le représentent est l'identité, c'est-à-dire que toutes les valeurs propres valent 1 et les vecteurs propres sont les vecteurs de la base canonique. Ainsi, pour représenter $(1 - \alpha)\%$ de la variance de ce bruit blanc, il faut considérer $(1 - \alpha)\%$ des vecteurs de la base canonique. Ceci explique la lenteur de la convergence des sommes cumulées des valeurs propres au delà du seuil $\Gamma(0)$. On peut dire que la décomposition

2.4. APPLICATION AUX MODÈLES GÉOSTATISTIQUES

de KL ne sera effective que lorsque la valeur de l'effet pépité p sera supérieure à α ; ainsi, dans ce cas, la part du bruit blanc pourra être négligée.

Conclusion

Nous avons pu observer dans cet étude l'apport potentiel de la décomposition de KL pour trois modèles géostatistiques couramment utilisés pour la description des propriétés pétrophysiques de réservoir, à savoir les modèles gaussiens munis des structures de covariance normale, exponentielle et sphérique. La méthodologie peut directement s'étendre à tout modèle gaussien de covariance continue.

Le principal intérêt offert par cette méthode est la réduction drastique de la dimension du problème de calage d'historique. En effet, prenons l'exemple du modèle de covariance normal, pour un champ en 2D carré de côté égal à 4 fois la portée pratique, discrétisé sur un maillage régulier comptant 10000 mailles. On a ainsi 10000 paramètres à caler, et donc à optimiser. Si l'on tronque la décomposition de K-L pour qu'elle soit représentative de 90% de la variance du processus représentant la perméabilité, il nous suffira de conditionner les 40 premières composantes de la décomposition, comme en atteste la figure (2.5).

2.4.2 Recherche d'un critère dynamique de choix du nombre de composantes

Problématique

Dans la section précédente, nous nous sommes attachés à déterminer le nombre de composantes nécessaires à reproduire un certain pourcentage de la variance du modèle considéré. Ce critère a été choisi afin de faciliter les comparaisons entre les différents modèles de covariance. Pour autant, on peut se demander si ce critère est adapté à la problématique des écoulements en milieux poreux, où les caractéristiques du milieu sont modélisées par un champ gaussien muni d'une certaine structure de covariance. En particulier, la question à laquelle nous allons tenter de répondre dans cette partie est : quelle part de la variance doit être employée pour reproduire raisonnablement les écoulements observés sur la réalisation de référence (non tronquée) ? Plus généralement, on va observer comment se comporte l'erreur commise par rapport à la simulation de référence pour des approximations à différents niveaux de troncature.

Nous allons pour ce faire employer le simulateur d'écoulement 3dsl [3ds, 2003], qui opère par lignes de courant, sur une durée totale de 10000 jours, pour simuler une période équivalente à la durée d'exploitation d'un réservoir. Le principe du simulateur d'écoulement à ligne de courant (streamline) est de transformer un problème d'écoulement en 2D ou 3D en une somme de problèmes en 1D, le long des lignes de courant, qui sont des trajectoires représentatives de l'écoulement global, déterminées par les conditions du réservoir. Nous représentons sur la figure 2.11 quelques lignes de courant, reliant un puits injecteur et un puits producteur, sur une carte de perméabilité.

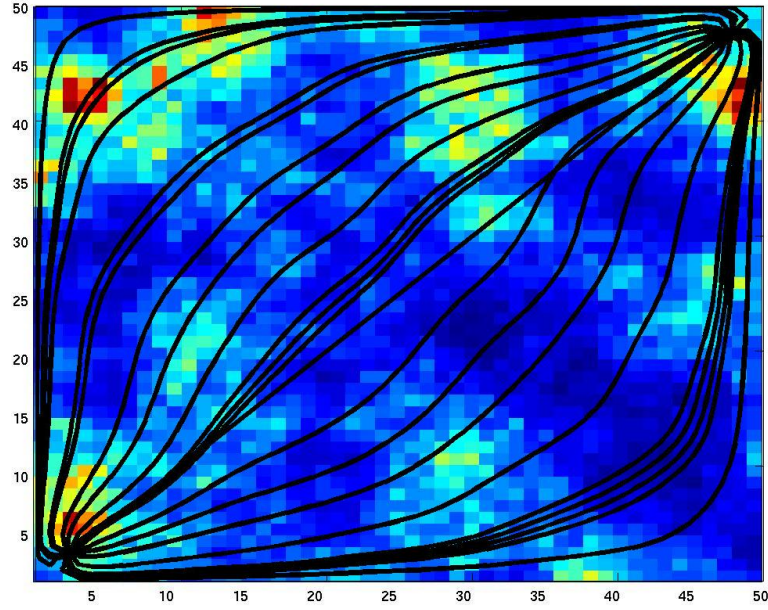


FIG. 2.11 – Lignes de courant reliant un puits injecteur et un puits producteur sur une carte de perméabilité.

Présentation du cas test

Nous avons effectué des simulations pour 100 réalisations d'un champ de perméabilité continu 2D dont les caractéristiques sont les suivantes :

- sa taille est de $2500 \times 2500 m^2$,
- il est discrétisé sur une grille régulière de $N = 50 \times 50$ blocs,
- son épaisseur est de $10 m$,
- son logarithme est un champ aléatoire gaussien Y de moyenne $\mu = 3.5$ et d'écart-type $\sigma = 0.5$, muni d'une structure de covariance sphérique (2.4.1, p. 43) de portée $600 m$ (1/4 de la taille du côté du champ),
- il est conditionné par les données de perméabilité aux puits, à savoir un injecteur situé à la position $(3, 3)$ de la grille, et un producteur situé à la position $(48, 48)$, de $90 md$,
- la porosité est supposée constante à 0.25 .

On a simulé 100 réalisations log-normales $(X_{(1)}, \dots, X_{(100)})$ de ce modèle et pour chaque réalisation, on a calculé les approximations dans la base de vecteurs propres pour les nombres de composantes suivants : $\{1, 2, \dots, 100, 105, \dots, 250, 260, \dots, 2490\}$, la réalisation complète comporte 2498 composantes, c'est-à-dire l'ensemble des blocs de la grille ôté des 2 blocs conditionnants. Cela représente un total de 35500 simulations. On notera l'approximation à n composantes de la réalisation i de la manière suivante : $\hat{X}_{(i)}^{(n)}$.

2.4. APPLICATION AUX MODÈLES GÉOSTATISTIQUES

Conditionnement Il est important de signaler que la matrice diagonalisée n'est pas exactement de la forme des matrices Γ vues en 2.4.1, p. 43. En effet, comme le champ est conditionné par la donnée de sa valeur en 2 noeuds i et j de la grille, on considérera la matrice de covariance conditionnelle Γ^c :

$$\Gamma_{N-2,N-2}^c = \Gamma_{N-2,N-2} - \Gamma_{N-2,2} \Gamma_{2,2}^{-1} \Gamma_{N-2,2}^t,$$

où : $\Gamma_{N-2,N-2}$ est la matrice de covariance entre les blocs non conditionnés,
 $\Gamma_{N-2,2}$ est la matrice de covariance entre les blocs non conditionnés et conditionnés,
 $\Gamma_{2,2}$ est la matrice de covariance entre les blocs conditionnés,
 $^{-1}, ^t$ sont respectivement l'opérateur inverse et l'opérateur transposé.

La moyenne conditionnelle du champ est donnée par :

$$\mu^c = \mu + \Gamma_{N-2,2}^t \Gamma_{2,2}^{-1} x_{i,j},$$

où $x_{i,j}$ est le vecteur des données de perméabilité en i et j .

On a alors $(X|X_i = x_i, X_j = x_j) \sim \mathcal{N}(\mu^c, \Gamma^c)$. Les valeurs propres et vecteurs propres calculés sont ici ceux de Γ^c .

Notre expérience simule la production d'huile sur une période de 10000 jours, soit environ 27 ans. Le milieu est supposé saturé en huile au temps 0 et la production est stimulée par injection d'eau. Le débit d'injection est constant à 5000 $m^3/jour$ et la pression au puits producteur est fixée à 200 *bars*. La valeur maximale du pas de temps a été fixée à 100 jours afin de minimiser le temps de simulation.

Nous représentons sur la figure 2.12 l'ensemble des 100 courbes de water cut correspondant aux 100 simulations de référence, c'est-à-dire générées avec l'ensemble des 2498 composantes, et la variance associée (en trait discontinu rouge). On s'assure ainsi de la bonne dispersion des courbes de water cut étudiées et des temps de percée associés. Le water cut mesure la part de l'eau dans la production de fluide et le temps de percée correspond , au moment à partir duquel l'on commence à produire de l'eau.

Résultats

Nous allons baser notre étude sur le comportement du water cut en fonction du temps. C'est une sortie fonctionnelle discrétisée sur le pas de temps. Plus précisément, on va regarder l'erreur relative ε commise par rapport à la réalisation géostatistique de référence, i.e non approximée. En notant, F l'opérateur d'écoulement qui à une réalisation X du champ de perméabilité associe le water cut, cette quantité s'exprime de la manière suivante : $\varepsilon = \frac{\|F(\hat{X}_{(i)}^{(n)}) - F(X_{(i)})\|^2}{F(\bar{X}_{(i)})^2}$, où $\|\cdot\|$ désigne la norme euclidienne et $\bar{\cdot}$, la moyenne. On examinera également l'erreur commise sur l'évaluation du temps de percée en fonction du nombre de composantes. Cette erreur est calculée comme la valeur absolue de l'écart entre les temps de percée de la réalisation de référence et son approximation à un ordre donné de troncature.

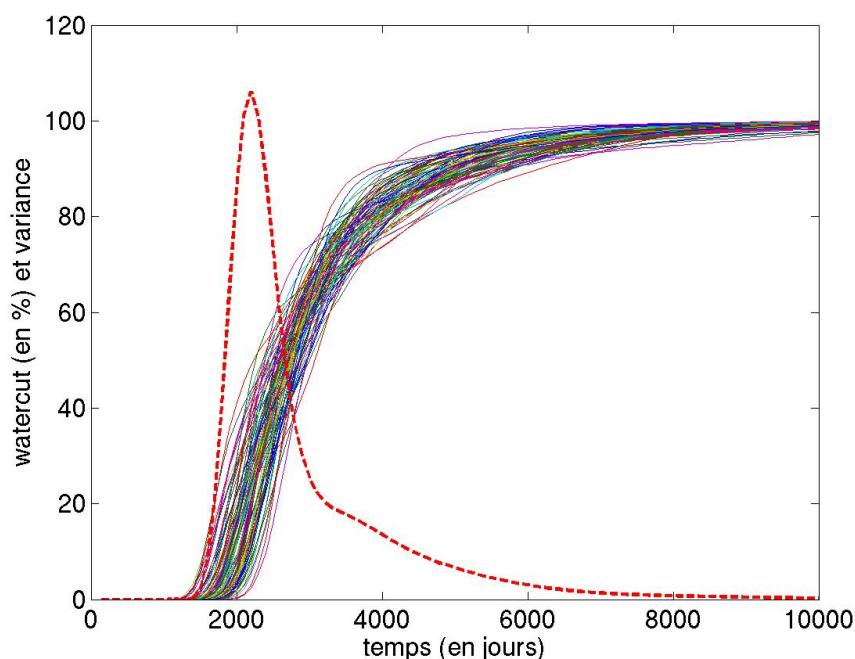


FIG. 2.12 – water cut des 100 simulations de référence et variance de l'échantillon (en trait discontinu rouge).

Dans un premier temps, nous allons observer précisément le comportement de ces deux réponses pour une unique réalisation aux différents ordres de troncature considérés. Dans un second temps, nous en regarderons le comportement moyen sur les 100 réalisations générées.

Etude d'une unique réalisation On représente respectivement sur les figures 2.13, 2.14 et 2.15, le water cut sur la période de temps simulée, correspondant à la réalisation complète et obtenu pour différents ordres de troncature, l'erreur relative en fonction du nombre de composantes de l'approximation (échelle logarithmique) et l'erreur commise vis-à-vis du temps de percée, également en fonction du nombre de composantes (échelle logarithmique).

Sur la figure 2.13, on peut voir le profil d'une courbe de water cut pour la réalisation complète ainsi que pour différents ordres de troncature. On peut d'ores et déjà noter que les réalisations tronquées semblent bien reproduire le comportement dynamique du réservoir. En effet, hormis la courbe de water cut correspondant à la réalisation avec une seule composante, les water cut des réalisations tronquées sont tous très proches de la courbe de référence (en pointillé rouge), correspondant à la réalisation complète. Nous allons maintenant pouvoir quantifier ce phénomène à la lecture des figures 2.14 et 2.15.

On peut voir sur la figure 2.14 le comportement de l'erreur relative sur le water cut, définie plus haut, en fonction du nombre de composantes employées pour décrire le champ de perméabilité. Conformément à ce que l'on pouvait attendre, la courbe pré-

2.4. APPLICATION AUX MODÈLES GÉOSTATISTIQUES

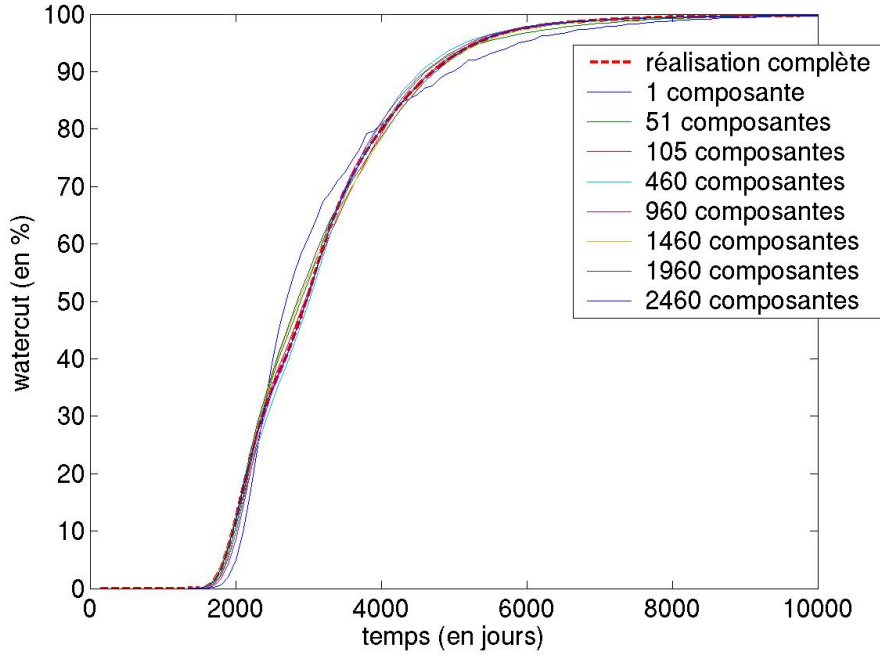


FIG. 2.13 – *Watercut obtenu pour la réalisation complète et à différents ordres de troncature.*

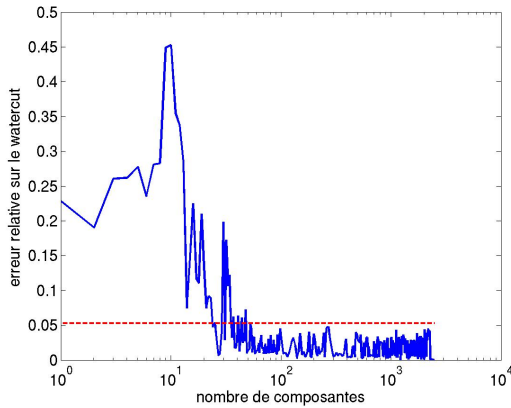


FIG. 2.14 – *Erreur relative sur le water cut en fonction du nombre de composantes de l'approximation .*

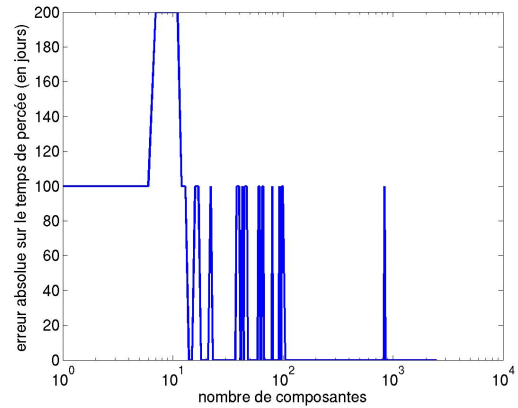


FIG. 2.15 – *Erreur absolue (en jours) commise vis-à-vis du temps de percée en fonction du nombre de composantes.*

sente une tendance décroissante. En outre, au-delà de 50 composantes environ, l'erreur commise est majorée par l'erreur “admissible” de 2% sur le water cut ($\frac{\|0.02 \times F(X_{(i)})\|^2}{F(\bar{X}_{(i)})^2}$). En ce qui concerne le temps de percée, l'erreur commise est faible par rapport au pas de temps employé (100 jours) et s'annule au-delà de 100 composantes hormis pour un seul ordre de troncature. La grande variabilité de ces résultats nous incite néanmoins à examiner les résultats agrégés sur les 100 simulations de références réalisées.

Etude d'un échantillon de 100 réalisations On représente respectivement sur les figures 2.16 et 2.17 l'erreur relative moyenne en fonction du nombre de composantes de l'approximation et l'erreur moyenne commise vis-à-vis du temps de percée également en fonction du nombre de composantes. On y trace également les intervalles de confiance à 95% de ces grandeurs construits par application du théorème central limite.

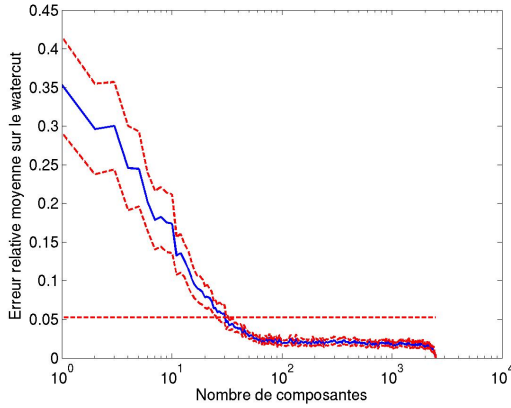


FIG. 2.16 – *Erreur relative moyenne sur le water cut en fonction du nombre de composantes de l'approximation et intervalle de confiance à 95% associé.*

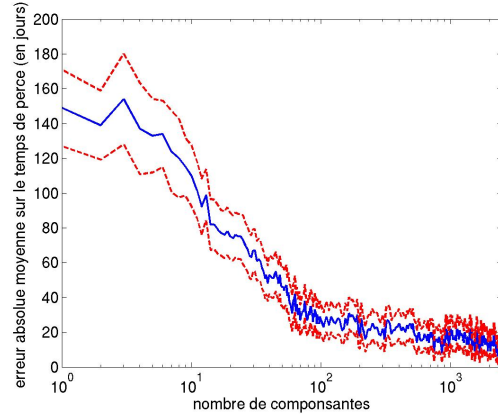


FIG. 2.17 – *Erreur absolue moyenne (en jours) commise vis-à-vis du temps de percée en fonction du nombre de composantes et intervalle de confiance à 95% associé.*

La figure 2.16 présente la moyenne, calculée sur l'ensemble des 100 réalisations de référence effectuées, de l'erreur relative sur le water cut (définie plus haut) en fonction du nombre de composantes de l'approximation et les intervalles de confiance à 95% associés. On peut voir que cette courbe possède un aspect bien moins erratique que celle de la figure 2.14. Elle présente un profil monotone avec une décroissance très rapide jusqu'au niveau des 50 composantes pour ensuite décroître très faiblement et enfin converger au-delà des 2000 composantes. On a reproduit sur ce graphe l'erreur admissible moyenne de 2%, définie plus haut, cette fois moyennée sur les 100 réalisations. On peut voir que la courbe de l'erreur moyenne passe sous ce seuil après 30 composantes puis tend à se stabiliser au-delà de 50. Dans le même temps, la variance de l'estimation se réduit considérablement lorsque l'on augmente le nombre de composantes comme le montre le resserrement des intervalles de confiance. On peut ainsi en déduire que pour cette application, un nombre réduit de composantes suffit à bien décrire le problème. Concernant l'erreur sur le temps de percée représentée sur la figure 2.17, elle présente un profil similaire à l'erreur relative, montrant néanmoins une convergence plus lente. On rappelle que ces résultats sont à mettre en regard du pas de temps employé dans la simulation d'écoulement, de 100 jours. Aussi, on peut voir qu'à partir de 10 composantes, l'erreur moyenne sur les temps de percée est inférieure à 1 pas de temps, à partir de 100, elle devient inférieure à 1/5 de pas de temps. De même que pour les résultats moyens sur le water cut, les intervalles de confiance se resserrent avec l'augmentation du nombre de composantes, ce qui assure une bonne stabilité à ces résultats. Une analyse plus fine de cette réponse pourrait être menée en diminuant celui-ci.

Conclusion

Ces résultats sont encourageants à plusieurs titres. Tout d'abord, on peut voir qu'un faible nombre de composantes suffit à honorer le critère dynamique, c'est-à-dire à reproduire de manière satisfaisante les résultats de la simulation d'écoulement. Ensuite, si l'on compare ces résultats à ceux de la section 2.4.1 précédente, on peut s'apercevoir que le nombre de composantes jugé ici suffisant est bien inférieur à celui nécessaire pour représenter 90% de la variance. Si l'on se reporte à la figure 2.5 page 46, on peut voir que 30 composantes correspondent à un taux de 70% de reproduction de la variance, pour le modèle de covariance employé. Ainsi, on peut affirmer que les faibles variations induites par les composantes d'ordre supérieur influent faiblement sur l'écoulement.

2.4.3 Discussion

On a montré dans cette partie le grand intérêt offert par la paramétrisation des champs aléatoires gaussiens par un développement tronqué de KL, pour la réduction de la dimension du problème inverse du calage d'historique. Les résultats de nature statistique de la partie 2.4.1 nous ont permis de quantifier le nombre de composantes nécessaires à reproduire approximativement la variabilité des modèles de champs aléatoires employés usuellement en géostatistique pétrolière. On a ensuite validé l'approche dans une problématique d'écoulement dans la partie 2.4.2. Les résultats sont encourageants et ouvrent la voie à une application rapide dans la problématique d'inversion.

Nous nous sommes volontairement restreints dans cette partie aux modèles gaussiens pour lesquels nous savons que les paramètres de la décomposition sont des gaussiennes indépendantes. On peut chercher à étendre la méthodologie dans le cadre plus large posé ici :

$$M = A(X),$$

où M représente le modèle d'étude, A un opérateur quelconque et X un processus gaussien muni de sa structure de covariance, donc décomposable selon la méthodologie présentée. On peut représenter de la sorte les modèles dits en faciès, pour lesquels A est un opérateur qui va tronquer les valeurs de la réalisation du processus selon un ou plusieurs seuils, pour faire apparaître les différents lithofaciès. Ce sera l'objet de la section suivante.

2.5 Modèles de gaussiennes seuillées

Les modèles de gaussiennes seuillées, ou modèles "en faciès", sont utilisés en ingénierie de réservoir pour représenter la répartition des différents lithofaciès dans le réservoir, auxquels sont ensuite associées des propriétés pétrophysiques. Ces modèles permettent de reproduire leurs proportions et leur distribution spatiale (inférées). On reprend dans cette section les notations des parties 2.2 et 2.3. Tout au long de cette section, le terme "troncature" se rapportera à la troncature du nombre de composantes dans la décomposition de KL d'un processus, tandis que le terme "seuillage" désignera l'opération qui consiste à générer les différents faciès à partir du processus continu sous-jacent.

2.5.1 Définitions

Un modèle de gaussienne seuillée est défini par la donnée d'un processus gaussien X_u indexé sur \mathbf{U} , muni d'une structure de covariance Γ stationnaire, et d'une ou plusieurs fonction de seuillage (S) définie sur \mathbf{U} . Le processus de gaussienne seuillée I ainsi défini prendra ses valeurs dans un ensemble fermé de \mathbb{N} . Dans la suite, on considérera X_u stationnaire. Le cas le plus simple étant celui d'une unique fonction S constante :

$$I_u(\omega) = \mathbb{1}_{\{S < X_u(\omega)\}}, \quad \omega \in \Omega,$$

I ainsi défini est d'espérance $\mu = 1 - G(S)$ où G représente la fonction de répartition gaussienne, centrée, de variance $\Gamma(0) = 1$, et de variance $\sigma^2 = G(S)(1 - G(S))$. On peut déduire sa covariance de la covariance de X de la manière suivante :

$$\begin{aligned} \text{cov}(I_u, I_{u+h}) &= \mathbb{E}(\mathbb{1}_{\{S < X_u\}} \mathbb{1}_{\{S < X_{u+h}\}}) - \mathbb{E}(\mathbb{1}_{\{S < X_u\}}) \mathbb{E}(\mathbb{1}_{\{S < X_{u+h}\}}) \\ &= \mathbb{E}(\mathbb{1}_{\{S < X_u\}} \mathbb{1}_{\{S < X_{u+h}\}}) - (1 - G(S))^2 \end{aligned}$$

Comme X_u est gaussien, on a :

$$\begin{aligned} X_{u+h} &= \frac{\Gamma(h)}{\Gamma(0)} X_u + \sqrt{\frac{\Gamma(0)^2 - \Gamma(h)^2}{G(0)}} z, \quad \text{où } z \sim \mathcal{N}(0, 1) \text{ et est indépendant de } X_u \\ &= \alpha_h X_u + \beta_h z \end{aligned}$$

d'où :

$$\begin{aligned} \mathbb{E}(\mathbb{1}_{\{S < X_u\}} \mathbb{1}_{\{S < X_{u+h}\}}) &= \mathbb{E}(\mathbb{1}_{\{S < X_u\}} \mathbb{1}_{\{S < \alpha_h X_u + \beta_h z\}}) \\ &= \int_S^{+\infty} g(x) \int_{\frac{S - \alpha_h x}{\beta_h}}^{+\infty} g(y) dy dx \\ &= \int_S^{+\infty} g(x) (1 - G(\frac{S - \alpha_h x}{\beta_h})) dx, \\ \text{où } g &\text{ désigne la densité de la loi normale centrée réduite.} \end{aligned}$$

On en déduit l'expression générale de la covariance de I :

$$Cov(I_u, I_{u+h}) = G(S)(1 - G(S)) - \int_S^{+\infty} g(x)G\left(\frac{S - \alpha_h x}{\beta_h}\right) dx, \quad (2.19)$$

$$\text{où } \alpha_h = \frac{\Gamma(h)}{\Gamma(0)} \text{ et } \beta_h = \sqrt{\frac{\Gamma(0)^2 - \Gamma(h)^2}{\Gamma(0)}}$$

Le terme intégral de (2.19) n'a pas de solution analytique ; il doit être calculé numériquement. Dans le cas d'une fonction de seuillage non constante, il suffit de remplacer S par $S(u)$ dans l'expression précédente ainsi que dans l'expression de la moyenne. Dans le cas où l'on a n fonctions de seuil S_1, \dots, S_n (constantes), les calculs sont plus complexes. On définit alors I^n de la manière suivante :

$$I_u^n(\omega) = \sum_{i=1}^n \mathbb{1}_{\{S_i < X_u(\omega)\}} = \sum_{i=1}^{n-1} i \mathbb{1}_{\{S_i < X_u(\omega) < S_{i+1}\}}, \quad u \in \mathbf{U}, \quad \omega \in \Omega,$$

d'espérance $\mu = n - \sum_{i=1}^n (1 - G(S_i))$ et de variance $\sigma^2 = \sum_{i=1}^{n-1} i^2 (G(S_{i+1}) - G(S_i)) - \mu^2$. La covariance prend alors l'expression suivante :

$$Cov(I_u^n, I_{u+h}^n) = \sum_{i,j} ij \int_{S_i}^{S_{i+1}} g(x) \left(G\left(\frac{S_{j+1} - \alpha_h x}{\beta_h}\right) - G\left(\frac{S_j - \alpha_h x}{\beta_h}\right) \right) dx,$$

$$\text{où } \alpha_h = \frac{\Gamma(h)}{\Gamma(0)} \text{ et } \beta_h = \sqrt{\frac{\Gamma(0)^2 - \Gamma(h)^2}{\Gamma(0)}}$$

Cette expression est encore valable dans le cas de fonctions de seuillage $\{S_i(u)\}$ non constantes, de la même façon que pour I . L'expression de la covariance peut paraître complexe, elle nous semble néanmoins plus simple à manier et à implémenter numériquement que la représentation en polynômes d'Hermite employée usuellement en géostatistique (voir par exemple [Emery, 2004] ou [Hu and HERESIM, 1995]).

On peut remarquer que la continuité de la covariance de la gaussienne seuillée découle directement de la continuité de la covariance sous-jacente. Par ailleurs, les calculs se déduisent de façon triviale dans le cas où X_u n'est pas stationnaire.

2.5.2 Choix du modèle de covariance pour la modélisation en faciès

Avant d'aborder les résultats numériques de l'application de la décomposition de KL aux modèles en faciès, il convient d'ouvrir une parenthèse à propos du choix de modèle de covariance (ou variogramme) utilisé dans ces modèles.

En général, en géostatistique pétrolière, les variogrammes employés pour la modélisation en faciès sont les modèles gaussien et sphérique. Une attention particulière est à porter sur ce dernier, notamment sur le comportement des trajectoires qu'il entraîne. En effet, ce variogramme étant linéaire à l'origine, ses trajectoires ne sont donc pas différentiables et donc très irrégulières. Plus précisément, en dimension 1, le comportement local des trajectoires est semblable à celui du mouvement brownien. En observant

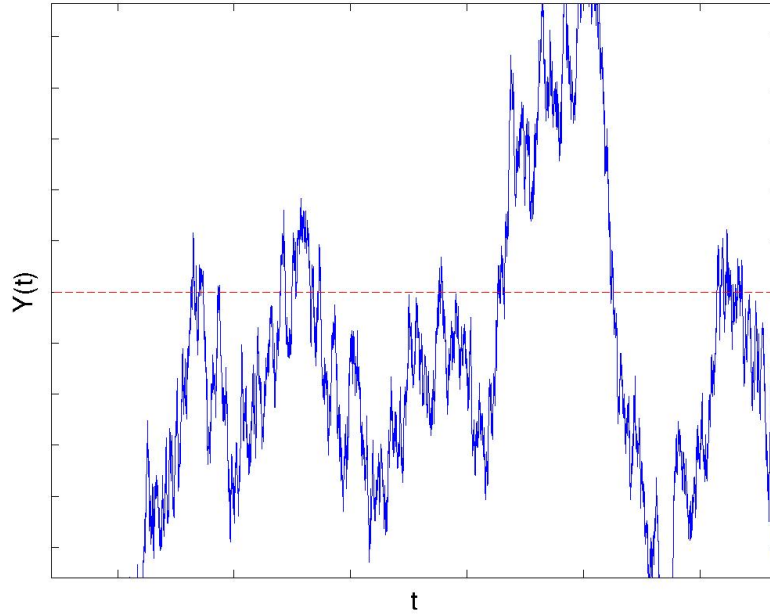


FIG. 2.18 – Comportement d’une trajectoire brownienne (pas de discrétisation de 0.0001) lors du passage d’un seuil quelconque.

une trajectoire brownienne au “passage” d’un seuil sur la figure 2.18, on se rend compte de ce que cela va entraîner au niveau des faciès. On voit bien que le nombre de “traversées” du seuil est considérable au niveau local. Il est à vrai dire infini, la discrétisation de la trajectoire entraînant nécessairement un nombre fini de passages du seuil. Cela se traduit dans le modèle en faciès par des oscillations très importantes entre les deux faciès à leur interface. Ces oscillations sont d’ailleurs d’autant plus nombreuses que le pas de discrétisation est fin.

Une manière de quantifier l’irrégularité d’une trajectoire au voisinage d’un seuil est donné par la formule de Rice, dont on donne l’énoncé en dimension 1 [Rice, 1945] :

Proposition 2.5.1 (Formule de Rice) *Soit X_u processus gaussien stationnaire sur $[0, U]$, de covariance $\Gamma(h)$. Pour w réel, on définit :*

$$N_u(w) = \#\{v \leq u, X_v = w\},$$

où $\#$ désigne le cardinal d’un ensemble. On a :

$$\mathbb{E}(N_u(w)) = ue^{-w^2/2} \frac{\sqrt{-\Gamma''(0)}}{\pi} \quad (2.20)$$

Cette quantité est donc finie si et seulement si la covariance (ou le variogramme) est deux fois dérivable en 0, ce qui est équivalent à une trajectoire X_t continûment différentiable (en moyenne quadratique). On trouve dans [Adler and Taylor, 2006] une généralisation de (2.20) aux dimensions supérieures. Il apparaît encore une fois que

la condition nécessaire et suffisante pour que l'espérance du nombre de franchissement d'un seuil soit localement finie est que la trajectoire du processus gaussien soit différentiable et donc que sa covariance soit deux fois différentiable en 0. Nous ne reproduisons pas ici le résultat car il fait appel à des notions de géométrie qui sortent largement du cadre de ce travail. Notons que dans le cas de la covariance sphérique (cf. 2.4.1 p. 43), cette espérance est infinie, ainsi, plus on raffine le modèle en faciès, plus ce comportement erratique à l'interface entre deux faciès apparaîtra. Bien que largement répandu dans la pratique, l'emploi de ce modèle semble ainsi inopportun dans le cas de gaussiennes seuillées.

On a ainsi mis en lumière un problème de spécification de la covariance pour les modèles en faciès. Dans un souci de cohérence des modèles géostatistiques indépendamment de leur échelle de précision, on préconisera donc dans la suite l'emploi de fonctions de covariance deux fois différentiables à l'origine. En effet, cette propriété assure un nombre fini de traversées du seuil à horizon fini. On donne ici la définition de la famille de covariances (notée ici C) dite de Matérn [Matérn, 1986] :

$$C(h) = \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\frac{h}{a}\right)^{\nu} K_{\nu}\left(\frac{h}{a}\right), \quad (2.21)$$

où a est un paramètre d'échelle et ν est un paramètre qui contrôle la régularité de la fonction, Γ représente ici la fonction gamma d'Euler, K_{ν} est la fonction de Bessel modifiée de seconde espèce. Les covariances exponentielle ($\nu = \frac{1}{2}$) et gaussienne ($\nu \rightarrow \infty$) en sont des cas particuliers.

L'emploi de ce modèle de covariance, bien que parfaitement adapté aux modèles en faciès pose plusieurs difficultés techniques :

- l'implémentation et le calcul de la fonction de Bessel modifiée de seconde espèce K_{ν} , particulièrement pour $\nu \notin \mathbb{N}$ (cf. [Press et al., 1992]) :
on dispose en effet d'une formule de récurrence pour calculer $K_n(x)$ en fonction de $K_{n-1}(x)$ et d'une approximation polynomiale de $K_0(x)$ (cf. [Abramowitz and Stegun, 1965]),
- On ne dispose pas de formule explicite pour la portée pratique $pp = C^{-1}(0.05) = \gamma^{-1}(0.95)$ de ce variogramme :
on trouve dans la littérature la valeur $pp = 2\sqrt{\nu}$ qui nous semble erronée. En effet, une étude expérimentale réalisée en faisant varier ν de 0 à 65 selon un pas de 1/8, pour $a = 1$ et un pas de discrétisation du variogramme de 10^{-3} sur l'intervalle $[0, 50]$ nous donne les résultats suivants :

$$pp \approx 3.4875\sqrt{\nu},$$

en optimisant le paramètre β_1 dans l'expression $\Gamma^{-1}(0.95) - \beta_1 \sqrt{\nu}$. Une meilleure approximation, au sens où la somme des carrés des écarts avec la "vraie" valeur ($pp = \arg \min\{\gamma(h) \geq 0.95\}$), est divisée par 4 par rapport à l'expression précédente, est :

$$pp \approx 3.6527\nu^{0.4874},$$

obtenue en optimisant β_2 et α dans $Ga^{-1}(0.95) - \beta_1 \nu^{\alpha}$. On retiendra donc cette approximation dans la pratique. Ces résultats sont illustrés sur la figure 2.19.

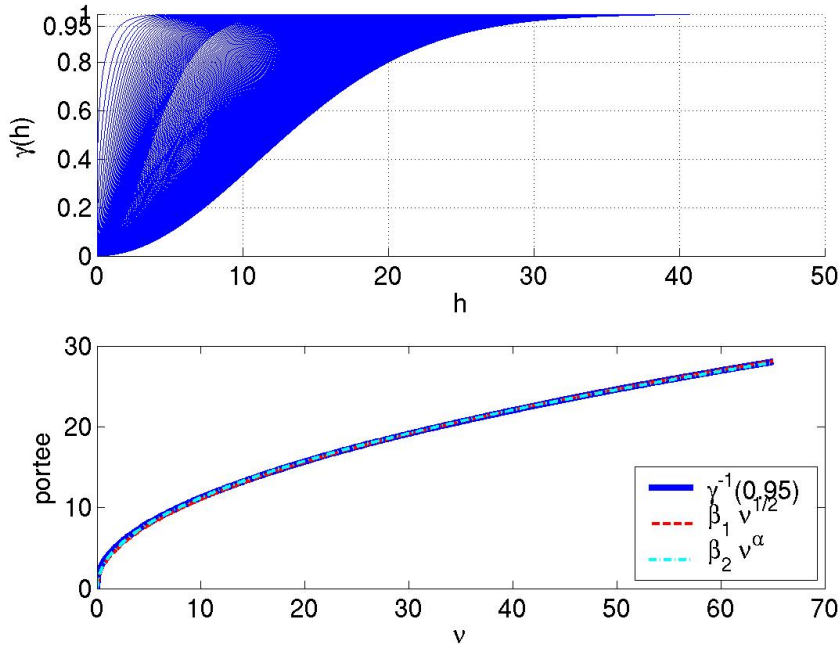


FIG. 2.19 – Représentation du variogramme de Matérn pour ν variant de 0 à 65 selon un pas de discrétisation de $1/8$ et portée pratique correspondante calculée et estimée .

Malgré ces difficultés, le variogramme de Matérn est le seul, parmi les variogrammes usuels, permettant de contrôler la régularité du processus gaussien associé et est donc d'un intérêt particulier pour les modèles en faciès, pour lesquels on a vu la nécessité de la différentiabilité en moyenne quadratique du processus gaussien sous-jacent. Dans la littérature ([Ecker, 2003], [Handcock and Stein, 1993]), on trouve que la condition nécessaire pour que ce variogramme soit 2 fois différentiable en 0 est $\nu \geq 2$.

2.5.3 Expériences numériques

A l'instar du travail réalisé sur les modèles continus de perméabilité, nous allons tester les performances numériques de la décomposition de KL sur les modèles en faciès, notamment l'influence de la dimension et de l'allongement de la portée.

Remarques préliminaires

Le principe d'application de la décomposition de KL aux modèles en faciès va être radicalement différent de la méthode proposée pour les modèles continus de perméabilité vus dans la section précédente. En vérité, deux méthodes différentes d'application de la décomposition de KL à ces modèles s'offrent à nous :

1. l'appliquer au processus gaussien sous-jacent, auquel cas on appliquera la fonction de seuillage à la version aproximée du processus gaussien.
2. l'appliquer au processus en faciès.

Dans la deuxième méthode envisagée, en raison de la nature intrinsèquement discontinue des processus considérés, il est impossible de les représenter dans une base incomplète de fonctions continues. La version tronquée du processus nécessitera un traitement que l'on nommera re-seuillage ou reconstitution pour représenter les faciès et la performance de cette opération déterminera le nombre de composantes retenues. En outre, le modèle en faciès n'est pas gaussien ; ses composantes dans le développement de KL ne sont pas gaussiennes et en conséquence, bien qu'orthogonales, c'est-à-dire non corrélées par construction, elles ne sont pas forcément indépendantes. Il convient alors d'inférer leur loi jointe afin de rendre possible leur simulation. Cette opération peut-être réalisée en estimant leur copule empirique (CE), i.e leur fonction de répartition empirique (fdre) jointe. On peut ensuite effectuer un test d'indépendance par comparaison avec les produits de fdre des variables considérées ([Genest and Remillard, 2004]). Une alternative à cette méthode, notamment si l'interdépendance d'un grand nombre de variables est avérée, ce qui entraînera des problèmes de stockage des CE, sera de construire un échantillon représentatif de la loi jointe, à partir duquel l'on peut ensuite estimer des quantiles et effectuer des simulations.

Présentation des modèles et modus operandi

Au vu de la discussion de la partie 2.5.2, nous étudierons les modèles de covariances Γ entraînant la nécessaire régularité des processus gaussiens associés, c'est-à-dire le modèle normal déjà vu dans la partie précédente et les modèles de Matérn (2.21). Le modèle en faciès considéré ici sera :

$$I_u = \mathbb{1}_{\{0.1 < X_u\}}, \quad u \in \mathbf{U}$$

Nous allons étudier les performances de la décomposition des modèles de gaussiennes seuillées au vu du pourcentage de points correctement reconstruits, en fonction du nombre de composantes retenues, *i.e.* :

$$P_n = 1 - \frac{1}{H} \sum_{i=1}^H |I_{u_i} - \hat{I}_{u_i}^n|,$$

où H représente le nombre de points u_i de la grille de discrétisation. Ce pourcentage ne peut pas être calculé théoriquement, nous allons donc adopter une démarche inférentielle, c'est-à-dire que nous allons générer un ensemble de M réalisations de I à partir duquel nous calculerons les projections sur le sous-espace vectoriel généré par les n premiers vecteurs propres de Γ , $n = 1..N$, et pour lesquels nous calculerons :

$$\hat{P}_n = 1 - \frac{1}{N} \sum_{j=1}^N \frac{1}{H} \sum_{i=1}^H |I_{j,u_i} - \hat{I}_{j,u_i}^n|.$$

Cette étude sera réalisée en comparant les deux méthodes évoquées plus haut :

1. la reconstitution issue de la décomposition de KL du processus gaussien sous-jacent :

$$\hat{I}_u^{N,1} = \mathbb{1}_{\{0.1 < \sum_{i=1}^N \xi_i \sqrt{\lambda_i} \phi_i(u)\}}$$

2. la reconstitution issue de la décomposition de KL du processus I_u lui-même :

$$\widehat{I}_u^{N,2} = \mathbb{1}_{\{0.5 < \mu + \sum_{i=1}^N \kappa_i \sqrt{\tau_i} \psi_i(u)\}}$$

Plus précisément, l'application de la deuxième méthode se déroule en deux étapes. On applique dans un premier temps la décomposition de KL au processus I_u , qui prend la forme suivante :

$$\widehat{I}_u^N = \mu + \sum_{i=1}^N \kappa_i \sqrt{\tau_i} \psi_i(u),$$

où μ est la moyenne du processus I_u , les κ_i sont les variables aléatoires de la décomposition de KL, les τ_i sont les valeurs propres de la covariance de I_u , et les $\psi_i(\cdot)$ sont les fonctions propres associés.

La seconde étape consiste à re-seuiller \widehat{I}_u^N à 0.5, afin d'obtenir un processus prenant pour valeur 0 ou 1. Nous calculons \widehat{P}_n en 1, 2 et 3 dimensions, pour des champs aléatoires discrétisés sur des grilles dont les dimensions sont respectivement 900, 30×30 et $10 \times 10 \times 10$. Pour estimer \widehat{P}_n , nous générons un ensemble de $M = 10000$ réalisations de champs aléatoires pour chaque variogramme et pour chaque dimension considérés. Chacune de ces réalisations est alors seuillée à 0.1.

Nous construisons alors les 10000 approximations $\widehat{I}_u^{N,1}$ en projetant les gaussiennes sous-jacentes sur leur base de vecteurs propres (ϕ_i). De même, nous construisons les $\widehat{I}_u^{N,2}$ en projetant les réalisations en faciès (seuillées) sur la base (ψ_i) de vecteurs propres de la covariance du modèle en faciès. Etant données, ces 2×10000 réalisations, on calcule alors $1 - \widehat{P}_n$, pour $n \in \{1, \dots, N_{max}\}$ et nous comparons les résultats graphiquement (figures 2.22 – 2.27, p. 64).

A titre d'exemple, nous représentons sur la figure 2.20 un champ 2D (60×60) gaussien de covariance normale de portée $pp = 15$ seuillé au niveau 0.1 et ses approximations, via la gaussienne sous-jacente, avec 20, 40, 60, 80 et 100 composantes.

Résultats

On représente en figure 2.21 les variogrammes considérés : les variogrammes des processus gaussiens sous-jacent et le variogramme du modèle en faciès correspondant calculé selon (2.19). Ce sont, comme la légende l'indique, les modèles de variogramme gaussien, et de Matérn avec $\nu = 2, 6$ avec des portées pratiques équivalentes.

Nous allons nous attacher maintenant à décrire et interpréter les différentes figures réalisées (2.22 – 2.27, p. 64). Sur ces figures, les courbes en trait plein correspondent aux valeurs de \widehat{P}_n pour la première méthode, où l'on décompose le processus gaussien sous-jacent, et les courbes en pointillés représentent \widehat{P}_n pour la seconde méthode.

On va observer dans un premier temps le comportement de \widehat{P}_n en 1 dimension (figures 2.22 et 2.23). La première chose que l'on remarque en comparant les courbes obtenues pour la décomposition de la trajectoire en faciès et de la trajectoire sous-jacente est que la régularité du processus sous-jacent domine le comportement de \widehat{P}_n . En particulier, pour $\nu \geq 4$ et pour le modèle de covariance normal, la décomposition de la trajectoire sous-jacente est plus performante que la décomposition de la trajectoire en faciès. En revanche, pour le modèle de covariance de Matérn avec $\nu = 2$, c'est cette

2.5. MODÈLES DE GAUSSIENNES SEUILLÉES

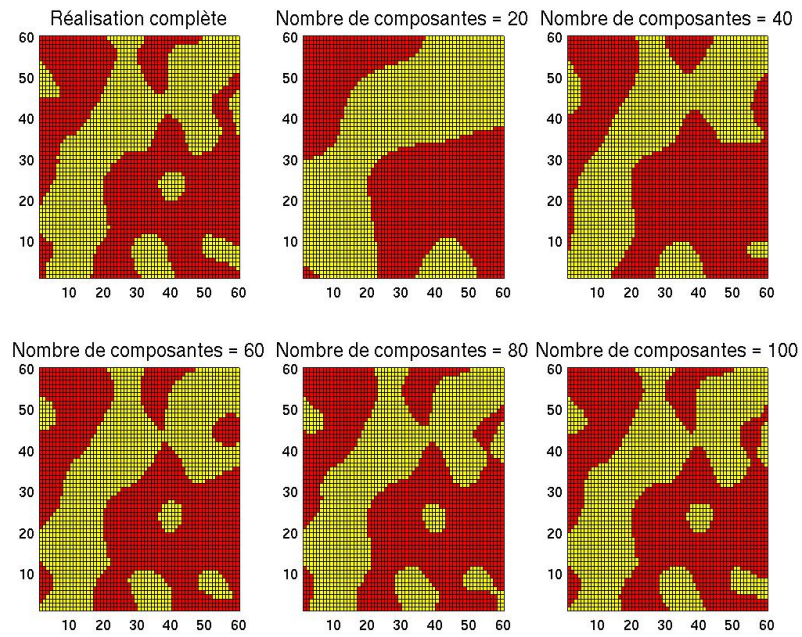


FIG. 2.20 – Exemple d'approximation d'un champ 2D en facies à différents ordres de troncature.

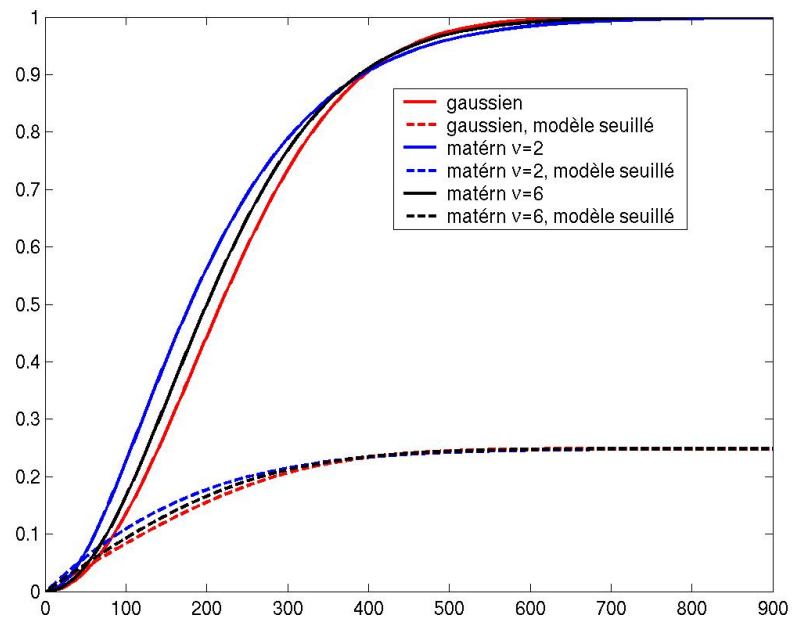


FIG. 2.21 – Variogrammes considérés et variogrammes du modèle en facies (seuil 0.1) correspondant.

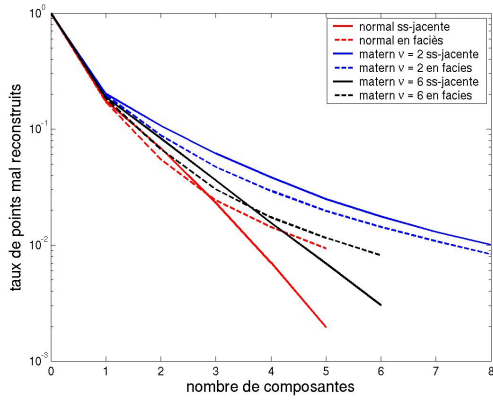


FIG. 2.22 – Taux de reconstitution erronée, cas 1D, $pp = H$.

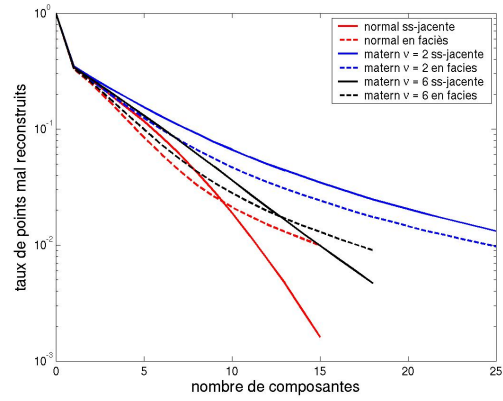


FIG. 2.23 – Taux de reconstitution erronée, cas 1D, $pp = H/4$.

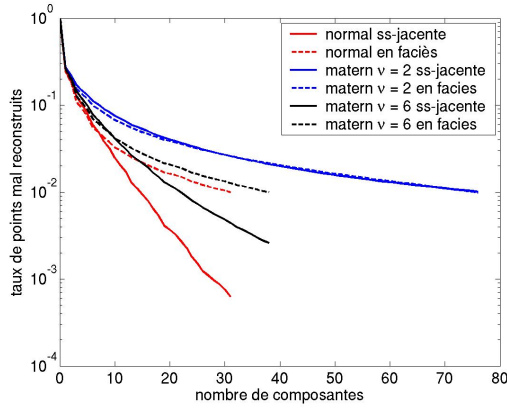


FIG. 2.24 – Taux de reconstitution erronée, cas 2D, $pp = H$.

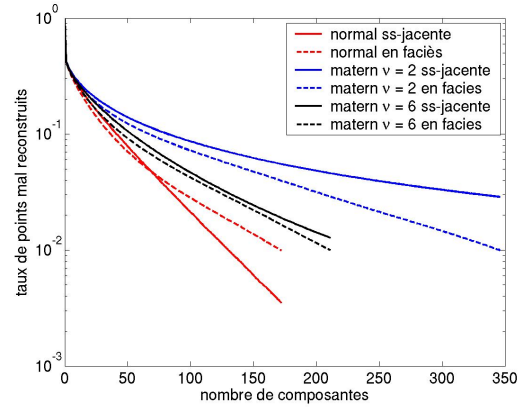


FIG. 2.25 – Taux de reconstitution erronée, cas 2D, $pp = H/4$.

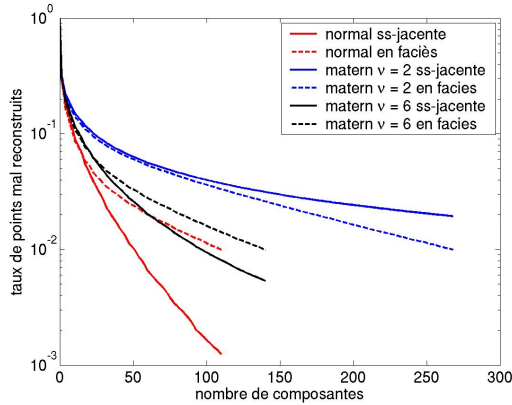


FIG. 2.26 – Taux de reconstitution erronée, cas 3D, $pp = H$.

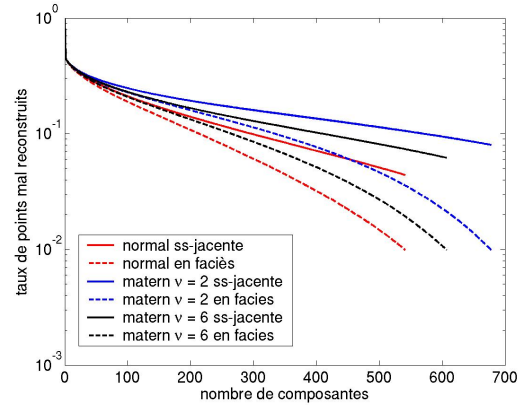


FIG. 2.27 – Taux de reconstitution erronée, cas 3D, $pp = H/4$.

dernière qui présente les meilleurs résultats.

Ce comportement semble différer en dimension 2 (figures 2.24 et 2.25). Il semble que

le niveau de régularité du processus sous-jacent doit être plus important, par rapport à la dimension 1, pour que la décomposition de la trajectoire sous-jacente soit plus performante. On voit d'autre part que le nombre de composantes nécessaires à une reconstitution au niveau 99% augmente considérablement.

De la même manière, lorsque l'on passe en dimension 3 (figures 2.26 et 2.27), les performances de la décomposition de la gaussienne sous-jacente s'amenuisent au point que dans le cas où la portée (pp) vaut le quart de la longueur d'un des cotés du domaine, elles deviennent inférieures à celles de la décomposition directe des réalisations en faciès. Cependant, dans le cas où pp est du même ordre de grandeur que la longueur des arêtes du domaine, la décomposition des champs sous-jacents donne de meilleurs résultats.

Il est difficile d'avancer une explication ou justification rigoureuse de ces résultats. Lorsque la dimension augmente, on peut suggérer que la géométrie du problème se complique, notamment au niveau des interfaces entre les deux faciès. En effet, en dimension 1, un point a 2 voisins directs, 4 en dimension 2 et 6 en dimension 3. Ceci peut éclairer en partie l'augmentation du nombre de composantes nécessaires à la reconstitution à 99%. Concernant le comportement de \hat{P}_n lorsque le rapport portée sur taille du domaine diminue, on peut dire qu'il augmente mécaniquement, dans le sens où lorsque l'on augmente la portée, cela revient à augmenter la taille du domaine, et ce phénomène est alors bien illustré par la proposition 2.5.1. On peut aussi s'interroger sur l'influence du pas de discrétisation, en particulier en ce qui concerne la figure 2.27. En effet, on ne dispose que de 10 points sur une arête donnée du cube de simulation et donc 3 points pour représenter le comportement entre l'origine et la portée. La nature régulière des champs correspondants est gommée par le très faible nombre de points de discrétisation. Ainsi, en notant, u_1, \dots, u_{10} les points de discrétisation sur une arête du cube considéré, on aura pour le modèle de covariance de matérn avec $\nu = 8$: $\Gamma(u_1) = 1$, $\Gamma(u_2) = 0.5744$ et $\Gamma(u_3) = 0.1331$; on perd ainsi le comportement très lisse de la covariance à l'origine.

On peut enfin pointer le cas du variogramme de Matérn où $\nu = 2$, qui diffère des autres modèles dans le sens où la décomposition du modèle en faciès présente de meilleures performances en terme de reconstitution que la décomposition de la gaussienne sous-jacente dans la plupart des cas de figures, notamment lorsque la dimension du modèle augmente (figures 2.24, 2.25, 2.26 et 2.27). Il conviendra par la suite d'inférer la loi jointe des composantes de la décomposition en faciès.

Ces résultats sont cependant sujets à caution, du fait que nous ne connaissons pas les propriétés asymptotiques de \hat{P}_n , notamment dans le cas où les composantes de la décomposition sont dépendantes les unes des autres. Si par exemple dix d'entre elles présentent des dépendances croisées, un échantillon de 10000 points paraît faible pour échantillonner une loi dans \mathbb{R}^{10} et l'estimation de P_n sera donc de mauvaise qualité. Une étude des dépendances croisées des composantes dans la décomposition du modèle en faciès est donc à mener. Néanmoins, au cours des différentes simulations effectuées, les résultats obtenus sur le calcul des \hat{P}_n ont fait montre d'une grande stabilité.

Etude de la loi des $\{\kappa_i\}_{i \in I}$ dans la décomposition du modèle en faciès

Comme on l'a dit en 2.5.3, les composantes $\{\kappa_i\}_{i \in I}$ dans la décomposition du modèle en faciès ne sont pas gaussiennes (cf. figure 2.28) et bien que non-corrélées, elles ne sont pas indépendantes. Elles sont par ailleurs centrées et réduites.

On va s'attacher dans ce paragraphe à étudier la structure de dépendance de ces variables. Pour ce faire, on va dans un premier temps tester l'indépendance des couples $(\kappa_i, \kappa_j)_{i \neq j}$, avec la statistique du ρ de Spearman [Spearman, 1904], basée sur les coefficients de corrélations des rangs. Cette statistique présente l'avantage, par rapport au coefficient de corrélation de Pearson, de pouvoir quantifier des dépendances non-linéaires ; elle ne détecte néanmoins que les dépendances monotones. Elle est facilement implémentable et calculable pour un grand nombre de couples de variables. Toutefois, elle ne peut quantifier des dépendances autres que monotones. En outre, les tests d'indépendances basés sur le ρ de Spearman sont valables quel que soit la distribution de κ_i et κ_j . Ses propriétés, notamment son comportement asymptotique, ont été étudiées en profondeur dans [Hotelling and Pabst, 1936].

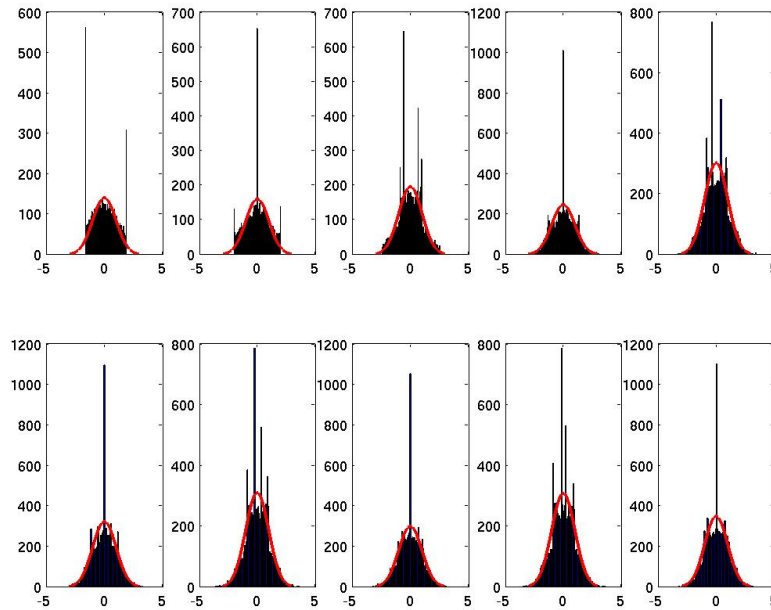


FIG. 2.28 – Histogrammes réalisés sur un échantillon de taille 10000 pour les 10 premières composantes de la décomposition du modèle en faciès (dimension 1) et densité gaussienne centrée réduite (en rouge).

Définition 2.5.1 Soit un échantillon (X_1, \dots, X_N) i.i.d. d'une variable aléatoire X de loi quelconque, et soit $(X_{i_1}, \dots, X_{i_N})$ tel que $X_{i_1} \leq X_{i_2} \leq \dots \leq X_{i_N}$. On définit la statistique de rang associé à cette échantillon de la manière suivante :

$$R^X = (i_1, \dots, i_N)$$

Pour un échantillon de taille N du couple $(\kappa_i, \kappa_j)_{i \neq j}$, le ρ de Spearman prend la forme suivante :

$$\rho_{i,j} = 1 - \frac{6}{N^2(N-1)} \sum_{k=1}^N (R_k^{\kappa_i} - R_k^{\kappa_j})^2, \quad (2.22)$$

où $R_k^{(i)}$, respectivement $R_k^{(j)}$, est la statistique de rang de κ_i , respectivement de κ_j . On a enfin, pour un grand échantillon, sous l'hypothèse d'indépendance, le résultat asymptotique suivant :

$$T = \rho \sqrt{N-1} \sim \mathcal{N}(0, 1), \quad (2.23)$$

Dans notre cas, où $N = 10000$, on construit donc facilement un test basé sur cette statistique dont la zone de rejet à 95% sera définie par $\{|T| > q_{0.975}\}$ où $q_{0.975}$ représente le quantile à 97.5% de la loi normale centrée réduite. Compte tenu du nombre élevé de composantes que l'on a pris en compte dans ces tests, on présente les résultats de manière agrégée, c'est-à-dire que les valeurs incluses dans le tableau 2.1 sont les moyennes des résultats des tests (0 ou 1) sur les C_{200}^2 combinaisons de deux variables. On présente ainsi dans le tableau 2.1 les valeurs agrégées des tests pour les modèles de variogrammes utilisés en 2.5.3, pour $(i, j) \in \{1..200\}^2, i \neq j$, en incluant en regard les valeurs agrégées des tests d'indépendance pour les composantes $(\xi)_{i=1..200}$ du processus gaussien sous-jacent.

Il semble que la structure de dépendance des $(\kappa_i)_{i=1..200}$ soit liée aux paramètres structurels du modèle considéré, en particulier sa dimension et le rapport portée pratique sur la longueur des arêtes du domaine. En effet, alors qu'en dimension 1 avec une portée pratique $pp = H$, les trois modèles de variogramme présentent un taux de rejet de l'hypothèse d'indépendance de l'ordre de 98%, en dimension 3, avec la même portée pratique, le taux de rejet pour chacun des trois modèles est du même ordre que les taux observés pour les ξ_i qui sont systématiquement aux alentours de 5%, le niveau de confiance du test ; on peut alors conclure à l'indépendance, au sens du ρ de Spearman de l'ensemble des couples de variables considérés.

Le modèle de variogramme considéré ne semble pas avoir d'impact sur les résultats de ces tests. En effet, pour chaque portée et dimension donnée, les 3 modèles donnent des résultats équivalents, hormis dans le cas 2D avec une portée pratique $pp = H$.

Pour autant, il faut se garder de conclusions hâtives. Observons précisément le cas en 2 dimensions, avec un variogramme de Matérn, $\nu = 2$ et une portée pratique $pp = H/4$. Les résultats du tableau 2.1 semblent montrer une indépendance des couples de variables (κ_i, κ_j) , $(i, j) \in \{1..200\}^2, i \neq j$. L'observation des histogrammes (voir figure 2.29) de ces variables nous a conduits à mener deux tests de normalité sur chacune des $(\kappa_i)_{i=1..200}$, à savoir le test de Kolmogorov-Smirnov basé sur le max de la différence entre la fonction de répartition empirique de l'échantillon (toujours de taille 10000), et le test de Jarque-Bera basé sur la comparaison des moments d'ordre 3 et 4 de l'échantillon avec ceux d'une gaussienne. Pour chacun de ces deux tests, le taux de rejet était aux alentours du niveau de confiance du test, c'est-à-dire de 5%. On serait ainsi tentés de conclure que les κ_i sont des v.a. gaussiennes indépendantes. Cette hypothèse ne résiste pas à la simulation. Si l'on compare visuellement des réalisations de ce modèle obtenues en tronquant la réalisation en faciès 200 composantes selon le principe du paragraphe

Variogramme dimension, portée	Taux de tests rejetés pour les $(\kappa_i)_{i=1..200}$	Taux de tests rejetés pour les $(\xi_i)_{i=1..200}$
normal 1D, $pp = H$	0.9877	0.0487
matern $\nu = 2$ 1D, $pp = H$	0.9742	0.0491
matern $\nu = 6$ 1D, $pp = H$	0.9754	0.0493
normal 1D, $pp = H/4$	0.0613	0.0510
matern $\nu = 2$ 1D, $pp = H/4$	0.0547	0.0510
matern $\nu = 6$ 1D, $pp = H/4$	0.0541	0.0510
normal 2D, $pp = H$	0.2777	0.0501
matern $\nu = 2$ 2D, $pp = H$	0.0602	0.0489
matern $\nu = 6$ 2D, $pp = H$	0.1400	0.0514
normal 2D, $pp = H/4$	0.0504	0.0516
matern $\nu = 2$ 2D, $pp = H/4$	0.0504	0.0514
matern $\nu = 6$ 2D, $pp = H/4$	0.0493	0.0513
normal 3D, $pp = H$	0.0518	0.0508
matern $\nu = 2$ 3D, $pp = H$	0.0509	0.0514
matern $\nu = 6$ 3D, $pp = H$	0.0507	0.0510
normal 3D, $pp = H/4$	0.0502	0.0501
matern $\nu = 2$ 3D, $pp = H/4$	0.0503	0.0495
matern $\nu = 6$ 3D, $pp = H/4$	0.0506	0.0501

TAB. 2.1 – Taux de rejet agrégés de l'hypothèse d'indépendance des couples de composantes parmi les 200 premières.

2.5.3 avec des réalisations simulées selon l'hypothèse précédente avec le même nombre de composantes, on se rend bien compte de son invalidité, comme en atteste la figure 2.30. En effet, alors que les figures de gauches exhibent des “plateaux” oscillants autour de 0 et 1 correspondant aux faciès (en rouge et bleu), celles de droites ne montrent pas de telles structures.

A noter que le comportement proche d'une gaussienne des κ_i semble se reproduire en dimension 2 et 3 lorsque $pp = H/4$.

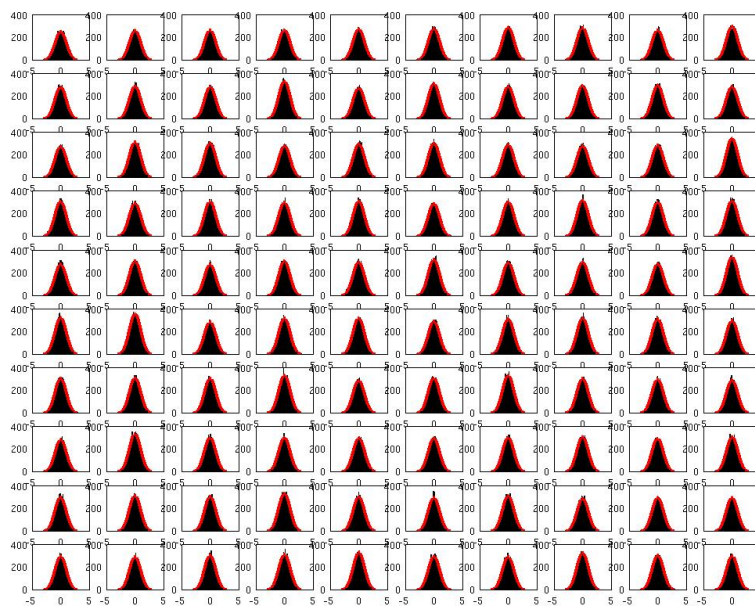


FIG. 2.29 – Histogrammes des 100 premières variables de la décomposition du modèle en faciès pour le variogramme de Matérn, $\nu = 2$ et une portée pratique $pp = H/4$.

Ainsi, bien que dans certains cas (voir tableau 2.1, page 68) les $(\kappa_i)_{i=1..200}$ soient indépendantes deux à deux, au sens du ρ de Spearman, leur structure de dépendance est vraisemblablement plus complexe. Ainsi, les n -uplets $n \geq 3$ peuvent présenter une structure de dépendance importante sans qu'elle soit pour autant détectable par les tests menés ci-dessus. Le seul outil capable d'inférer ces lois jointes est le copule, une généralisation de la fonction de répartition empirique aux n -uplets de variables. Son estimation pour un grand nombre de variables est néanmoins délicate pour deux raisons : la nécessité d'inférer une loi sur un espace de très grande dimension implique de disposer d'un très grand échantillon (convergence très lente de l'estimateur) présentant de bonnes propriétés de répartition et le problème du stockage induit par une estimation non-paramétrique d'une loi multi-dimensionnelle. En effet, si l'on considère que 1000 points suffisent pour avoir une estimation satisfaisante d'une loi monodimensionnelle, il faut alors 1000^{100} points pour estimer la loi jointe des 100 premières variables. Ce phénomène est une bonne illustration de la malédiction de la dimensionnalité (*curse of*

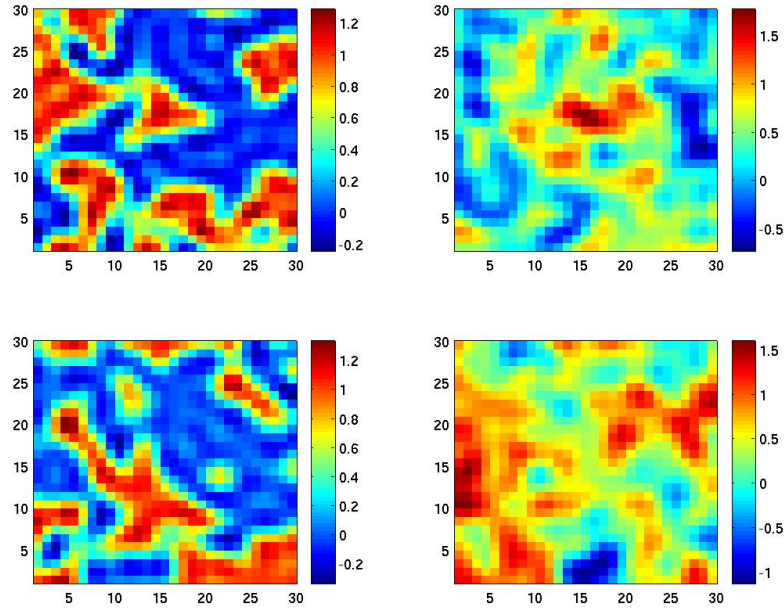


FIG. 2.30 – Réalisations du modèle en faciès approximé pour le variogramme de Matérn, $\nu = 2$ et une portée pratique $pp = H/4$. À gauche, les réalisations obtenues en tronquant une réalisations en faciès, à droite obtenues en simulant selon l'hypothèse de v.a. gaussiennes indépendantes .

dimensionality), problème récurrent en statistique lorsque la dimension du problème est importante.

Devant l'impossibilité d'inférer la loi jointe des (κ_i) par leur copule empirique, deux solutions restent possibles pour simuler directement les versions approximées des modèles en faciès :

- à l'image de ce qui a été fait dans l'étude menée ici, on pourrait créer un échantillon "représentatif" de réalisations de ces modèles,
- ou alors, on peut se borner à décomposer la gaussienne sous-jacente, bien que cette méthode soit sous-optimale dans certains cas, au vu des résultats de la partie 2.5.3.

Conclusion

La dernière solution proposée, c'est-à-dire décomposer le champ gaussien sous-jacent au modèle en faciès, semble la plus raisonnable et la plus simple à mettre en oeuvre. Il est en effet difficile de quantifier la représentativité d'un échantillon dans un espace de très grande dimension.

Cette méthodologie peut d'ailleurs s'étendre sans effort aux modèles pluri-gaussiens (voir [Armstrong et al., 2001]) qui se présentent de la manière suivante : étant donnés deux champs aléatoires gaussiens Y_1 et Y_2 , on construit deux nouveaux champs X_1 et

X_2 par combinaison linéaire :

$$\begin{aligned} X_1 &= Y_1 \\ X_2 &= \rho Y_1 + \sqrt{1 - \rho^2} Y_2, \end{aligned}$$

où ρ correspond à un coefficient de corrélation entre X_1 et X_2 .

Les relations entre les différents faciès peuvent être obtenues expérimentalement, par exemple à partir d'observations. L'agencement final des différentes relations sera un diagramme de faciès dans l'espace défini par les deux gaussiennes X_1 et X_2 . La méthode des pluri-gaussiennes tronquées apparaît ainsi comme une extension de la méthode des gaussiennes tronquées, permettant un libre agencement des faciès.

On appliquera la décomposition de KL à Y_1 et Y_2 pour réduire le nombre de paramètres du modèle.

2.5.4 Application aux écoulements

Problématique

Dans la section 2.5.3 précédente, nous nous sommes attachés à déterminer le nombre de composantes nécessaires à reproduire un certain pourcentage de reconstitution de la carte de répartition des faciès. Ce critère a été choisi afin de faciliter les comparaisons entre les différents modèles de covariance. Pour autant, on peut se demander si ce critère est adapté à la problématique des écoulements en milieux poreux. En effet, les écoulements seront facilités ou entravés selon le faciès rencontré ; à ce titre, on peut penser que des caractéristiques géométriques de la répartition des faciès, notamment la connexité des différents faciès, joueront un rôle important. A l'image de ce qui a été réalisé en 2.4.2 p. 49, on va observer comment se comporte l'erreur commise par rapport à la simulation de référence (non tronquée) pour des approximations à différents niveaux de troncature.

Nous allons pour ce faire employer le simulateur d'écoulement 3dsl [3ds, 2003], noté F , qui opère par lignes de courant sur une durée totale de 10000 jours, pour simuler une période équivalente à la durée d'exploitation d'un réservoir.

Présentation du cas test

Nous avons effectué des simulations pour 100 réalisations d'un modèle $2D$ en faciès dont les caractéristiques sont les suivantes :

- sa taille est de $2500 \times 2500 m^2$,
- il est discrétisé sur une grille régulière de $N = 50 \times 50$ blocs,
- son épaisseur est de $10 m$,
- la gaussienne sous-jacente est un champ aléatoire gaussien Y de moyenne $\mu = 0$ et d'écart-type $\sigma = 0.5$, muni d'une structure de covariance sphérique (2.4.1, p. 43) de portée $600 m$ ($1/4$ de la taille du côté du champ),
- il est composé de 2 faciès A et B , à hauteur de 30% de faciès A de perméabilité $300 md$ et 70% de faciès B de perméabilité $50 md$,
- il est muni d'un injecteur situé à la position $(3, 3)$ de la grille, et un producteur situé à la position $(48, 48)$,
- la porosité est supposée constante à 0.25.

On a simulé 100 réalisations en faciès $(X_{(1)}, \dots, X_{(100)})$ de ce modèle et pour chaque réalisation, on a calculé les cartes de faciès correspondant aux approximations de la gaussienne sous-jacente dans la base de vecteurs propres pour les nombres de composantes suivants : $\{1, 2, \dots, 100, 110, \dots, 250, 300, \dots, 2450\}$, la réalisation complète comporte 2500 composantes, c'est-à-dire l'ensemble des blocs de la grille. Cela représente un total de 15600 simulations. On notera l'approximation à n composantes de la réalisation i de la manière suivante : $\hat{X}_{(i)}^{(n)}$.

Résultats

De la même manière que dans le paragraphe 2.4.2 p. 49, les sorties observées seront l'erreur relative ($\varepsilon = \frac{\|F(\hat{X}_{(i)}^{(n)}) - F(X_{(i)})\|^2}{F(\bar{X}_{(i)})^2}$) commise sur le water cut et l'erreur commise sur l'évaluation du temps de percée, en fonction du nombre de composantes.

Dans un premier temps, nous allons observer précisément le comportement de ces deux réponses pour une unique réalisation aux différents ordres de troncature considérés. Dans un second temps, nous en regarderons le comportement moyen sur les 100 réalisations générées.

Etude d'une unique réalisation On représente sur les figures 2.31 et 2.32, l'erreur relative en fonction du nombre de composantes de l'approximation (échelle logarithmique) et l'erreur commise vis-à-vis du temps de percée également en fonction du nombre de composantes (échelle logarithmique).

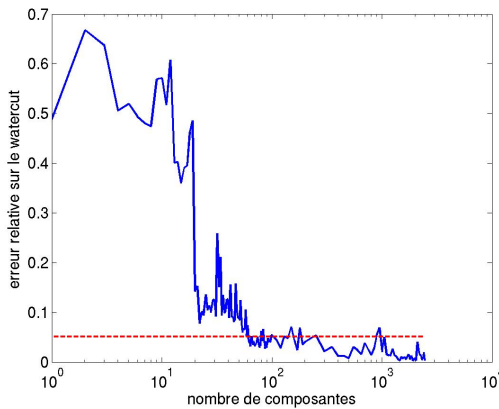


FIG. 2.31 – Erreur relative sur le water cut en fonction du nombre de composantes de l'approximation (réalisation 1).

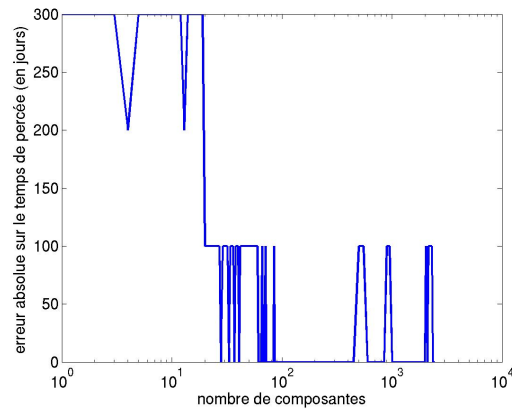


FIG. 2.32 – Erreur absolue (en jours) commise vis-à-vis du temps de percée en fonction du nombre de composantes (réalisation 1).

On peut voir sur la figure 2.31 le comportement de l'erreur relative sur le water cut, en fonction du nombre de composantes employées pour décrire la carte des faciès. Conformément à ce que l'on pouvait attendre, la courbe présente une tendance décroissante, malgré un comportement très erratique pour un nombre de composantes inférieur à 50. En outre, au-delà de 70 composantes environ, l'erreur commise ε est majorée, par

l'erreur "admissible" de 2% sur le water cut ($\frac{\|0.02 \times F(X_{(i)})\|^2}{F(\bar{X}_{(i)})^2}$). En ce qui concerne le temps de percée, l'erreur commise est faible par rapport au pas de temps employé (100 jours) et s'annule au-delà de 100 composantes hormis pour quelques réalisations.

A la différence du cas continu, on peut voir facilement sur les cartes de faciès les changements dans la géométrie (formes, connexité) occasionnés par l'ajout d'une composante et les relier au comportement de l'erreur en fonction du nombre de composantes. A ce titre, nous allons décrire précisément la réalisation 67 qui présente les plus fortes variations de l'erreur de l'échantillon. On représente sur les figures 2.33 et 2.34 les profils de l'erreur sur le water cut et le temps de percée en fonction du nombre de composantes de l'approximation.

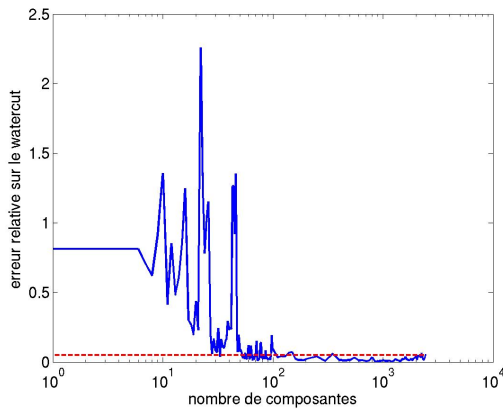


FIG. 2.33 – *Erreur relative sur le water cut en fonction du nombre de composantes de l'approximation (réalisation 67).*

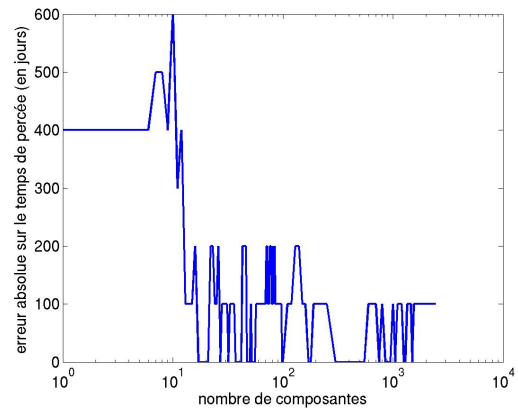


FIG. 2.34 – *Erreur absolue (en jours) commise vis-à-vis du temps de percée en fonction du nombre de composantes (réalisation 67).*

La figure 2.33 présente un comportement de l'erreur fortement erratique pour un nombre de composantes inférieur à 100. En particulier, on peut voir un pic de l'erreur pour 22 composantes alors que cette erreur est beaucoup plus faible pour 21 et 23 composantes, avec des valeurs de l'erreur de 2.2593, 0.2284 et 1.2343 respectivement. De même, l'erreur décroît drastiquement en passant de 46 à 47 puis 48 composantes avec des valeurs respectives de 1.3530, 0.1821 et 0.0749. Nous représentons sur la figure 2.35 les cartes de faciès correspondant à ces différents nombres de composantes, ainsi que la carte de référence, c'est-à-dire la représentation complète à 2500 composantes et l'approximation à 110 composantes, pour laquelle l'erreur vaut 0.0328. Sur ces cartes, le rouge correspond au faciès *A* dont la perméabilité vaut 300 *md*, le bleu au faciès *B* moins perméable avec 50 *md*. Nous représentons également sur la figure 2.36 les tracés des water cut correspondant à ces différentes approximations. Au vu des cartes correspondant à 21 et 22 composantes, on peut penser que l'augmentation de la valeur de l'erreur est due à l'apparition d'une connexion entre les deux blocs de faciès *A*, très perméable, situés près du centre de la carte, ainsi qu'à l'apparition d'un petit bloc de faciès *A* en bas à droite. Celui-ci disparaît lorsque l'on considère 23 composantes et la connexion entre les deux blocs centraux s'affine, ce qui semble expliquer le gain obtenu pour ce nombre de composantes.

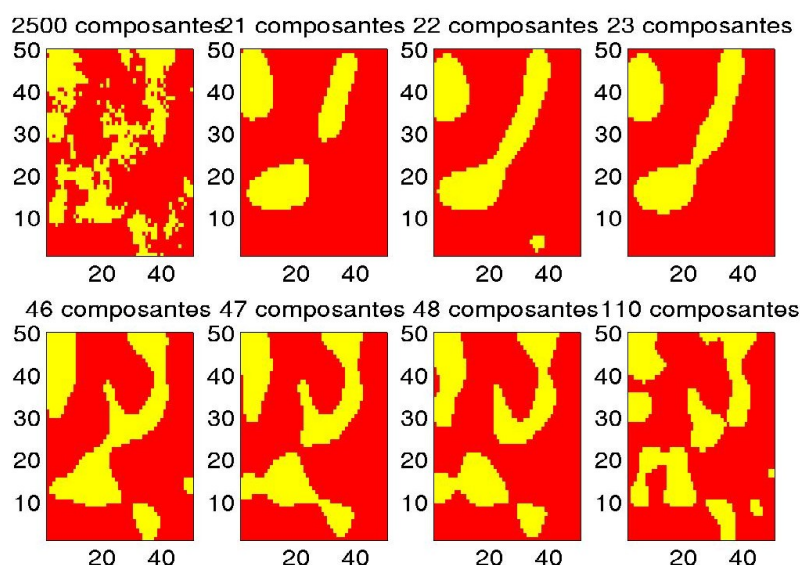


FIG. 2.35 – Cartes de faciès correspondant à différents nombres de composantes dans la décomposition (réalisation 67).

En ce qui concerne la baisse drastique de la valeur de l'erreur en passant de 46 à 47 puis à 48 composantes, elle semble là aussi s'expliquer par des modifications des connexions entre les grands blocs de faciès A . La carte correspondant à 46 composantes exhibe une connexion entre les 2 blocs vus dans la description ci-dessus. Cette connexion disparaît pour les cartes 47 et 48. En revanche, le passage de 46 à 47 composantes entraîne une connexion entre les deux blocs de faciès A situés en bas de ces cartes. Celle-ci disparaît pour 48 composantes, entraînant une diminution de l'erreur.

Il est important de comparer à chaque fois ces modifications de la géométrie de la carte avec la carte de référence et de les lier avec les courbes de water cut respectives. Dans les différentes cartes détaillées ci-dessus, nous avons identifié deux principaux problèmes. Le premier, que l'on peut qualifier de connexion centrale, a été discuté dans chacun des paragraphes précédents. Le second, que nous nommerons connexion basse, intervient uniquement dans le second. Le cas de la connexion centrale peut paraître paradoxal : c'est lorsqu'elle disparaît en modifiant le nombre de composantes que l'on observe une diminution de la valeur de l'erreur, alors qu'elle existe dans la carte de référence. Cependant, la forme qu'elle prend est très fine dans la carte de référence ainsi, le comportement dynamique sera plus proche dans cette région de celui observé sans la connexion qu'avec une connexion épaisse. En effet si l'on se rapporte à la figure 2.36, dans ces cas (22, 23 et 46 composantes) ce phénomène se traduit par une diminution du temps de percée fortement préjudiciable à l'erreur sur le water cut. Le cas de la connexion basse est plus direct. Cette connexion n'existe pas plus dans la carte de référence que dans la carte correspondant à 48 composantes qui présente de bons résultats.

Enfin, nous pouvons clore cette description en remarquant que la carte construite à

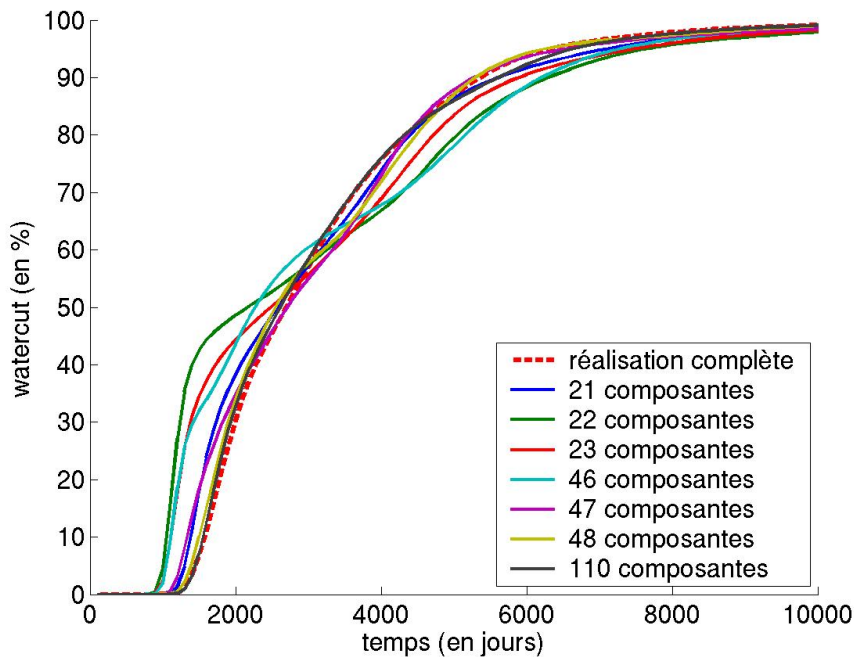


FIG. 2.36 – Courbes de water cut correspondant à différents nombres de composantes dans la décomposition (réalisation 67).

partir des 110 premières composantes reproduit raisonnablement bien la simulation de référence, notamment les grands blocs de faciès *A* et leur connexions, et surtout son comportement dynamique, avec une erreur de 0.0328 et une très bonne reproduction du temps de percée comme le montre la figure 2.36.

A l'image des résultats obtenus en 2.4.2 p. 49, la grande variabilité des erreurs commises sur le water cut et le temps de percée nous incite à examiner les résultats agrégés sur les 100 simulations de références réalisées afin d'en extraire les principales tendances.

Etude d'un échantillon de 100 réalisations On représente respectivement sur les figures 2.37 et 2.38 l'erreur relative moyenne en fonction du nombre de composantes de l'approximation et l'erreur moyenne commise vis-à-vis du temps de percée également en fonction du nombre de composantes. On y trace également les intervalles de confiance à 95% de ces grandeurs construits par application du théorème central limite.

La figure 2.37 présente la moyenne, calculée sur l'ensemble des 100 réalisations de références effectuées, de l'erreur relative sur le water cut (définie plus haut) en fonction du nombre de composantes de l'approximation et les intervalles de confiance à 95% associés. On peut s'apercevoir que cette courbe possède un aspect bien moins erratique que celle des figures 2.31 ou 2.33. Elle présente un profil monotone avec une décroissance très rapide jusqu'au niveau des 100 composantes pour ensuite décroître très faiblement et enfin converger au-delà des 2000 composantes. On a reproduit sur ce graphe l'erreur admissible moyenne de 2%, définie plus haut, cette fois moyennée sur les 100 réalisations. On peut voir que la courbe de l'erreur moyenne passe sous ce seuil après un peu

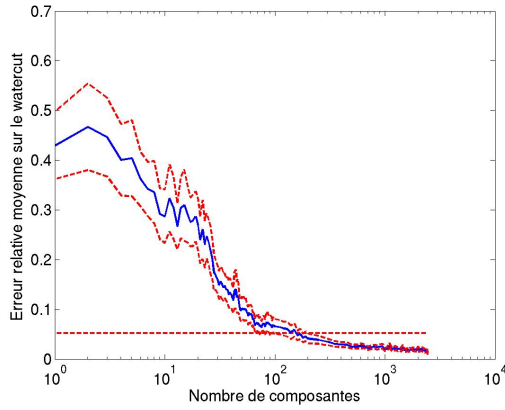


FIG. 2.37 – Erreur relative moyenne sur le water cut en fonction du nombre de composantes de l'approximation et intervalle de confiance à 95% associé.

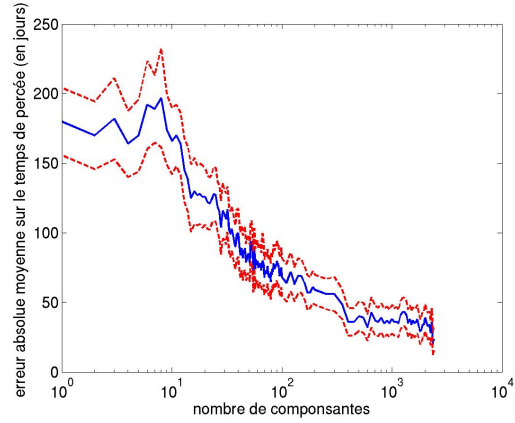


FIG. 2.38 – Erreur absolue moyenne (en jours) commise vis-à-vis du temps de percée en fonction du nombre de composantes et intervalle de confiance à 95% associé.

plus de 100 composantes. Dans le même temps, la variance de l'estimation se réduit considérablement lorsque l'on augmente le nombre de composantes comme le montre le resserrement des intervalles de confiance. On peut ainsi en déduire que pour cette application, un nombre réduit de composantes suffit à bien décrire le problème. En outre, la borne supérieure de l'intervalle de confiance passe sous le seuil de l'erreur admissible aux alentours des 200 composantes. Si l'on compare ces résultats avec ceux obtenus lors de la modélisation d'un champ de perméabilité continu (2.4.2 p. 49), on peut voir que les décroissances aussi bien de l'erreur moyenne que de sa variance sont ici moins rapides. Par conséquent, un nombre plus élevé de composantes dans l'approximation sera nécessaire à la description du champ en faciès.

Concernant l'erreur sur le temps de percée représentée sur la figure 2.38, elle présente un profil similaire à l'erreur relative, montrant néanmoins une convergence plus lente. On rappelle que ces résultats sont à mettre en regard du pas de temps employé dans la simulation d'écoulement, de 100 jours. Aussi, on peut voir qu'à partir de 40 composantes, l'erreur moyenne sur les temps de percée est inférieure à 1 pas de temps, à partir de 200, elle devient inférieure à 1/2 pas de temps. De même que pour les résultats moyens sur le water cut, les intervalles de confiance se resserrent avec l'augmentation du nombre de composantes, ce qui assure une bonne stabilité à ces résultats. Une analyse plus fine de cette réponse pourrait être menée en diminuant celui-ci, moyennant une augmentation importante des temps de calcul.

Conclusion

Ces résultats sont encourageants à plusieurs titres. Tout d'abord, on peut voir qu'un faible nombre de composantes, de 100 à 200 pour ce cas, suffit à honorer le critère dynamique, c'est-à-dire à reproduire de manière satisfaisante les résultats de la simulation d'écoulement. Ensuite, si l'on compare ces résultats à ceux de la section 2.5.3 précédente, on peut s'apercevoir que le nombre de composantes jugé ici suffisant est bien inférieur à celui nécessaire pour reconstituer 99% des points de la carte, qui fut le

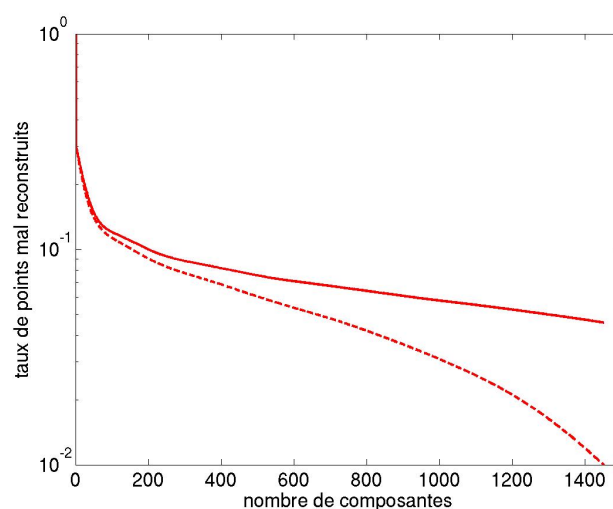


FIG. 2.39 – *Taux de reconstitution erronée pour le cas étudié.*

critère étudié précédemment. Nous avons calculé l'indicateur \hat{P}_n pour le cas étudié dans cette section (cf. 2.5.4) et représentons son comportement sur la figure 2.39. Comme précédemment, la courbe en trait plein correspond à la décomposition de la gaussienne sous-jacente et la courbe en tiret à la décomposition de la réalisation en faciès. En observant cette figure, on peut s'apercevoir que le nombre de composantes validées dans l'étude du comportement dynamique correspond à une reconstitution d'environ 90%. Si l'on met ce nombre en parallèle avec les résultats obtenus en 2.5.3, on obtient un indicateur du nombre de composantes nécessaires à reproduire le comportement dynamique pour un modèle géostatistique de répartition des faciès donné.

2.6 Autres applications de la décomposition de Karhunen-Loève

Nous présentons dans cette section quelques applications et considération pratiques portant sur l'application décomposition KL en ingénierie de réservoir. Pour commencer, nous allons voir comment il peut fournir un indicateur du nombre de réalisations nécessaires dans la boucle d'optimisation basée sur la déformation graduelle, employée à l'IFP. Dans un second temps, on s'attachera à décrire quelques aspects pratiques en vue d'une application à des modèles de réservoir réalistes. Enfin, nous évoquerons les modèles multipoints, déjà présentés en introduction, pour la paramétrisation desquels nous proposerons quelques idées, puis nous mentionnerons l'approche de résolution des équations d'écoulement dans un cadre stochastique.

2.6.1 Karhunen-Loève et déformation graduelle

Le procédé de calage employé actuellement à l'IFP est basé sur une paramétrisation permettant de déformer les réalisations d'un modèle stochastique donné tout en conservant leur structure spatiale. On le présente ici dans le cas de processus gaussiens. Étant donné deux fonctions aléatoires suivant la loi d'un modèle géostatistique X , processus gaussien centré de covariance Γ , X_1 et X_2 , on définit la déformation graduelle [Hu, 2000] de ces deux fonctions par :

$$X(t) = \cos(t)X_1 + \sin(t)X_2$$

On le généralise aisément à n fonctions de la manière suivante :

$$X(t_1, \dots, t_n) = \prod_{i=1}^{n-1} \cos(t_i)X_1 + \sum_{i=1}^{n-1} \sin(t_i) \prod_{j=i+1}^{n-1} \cos(t_j)X_{i+1}, \quad (2.24)$$

Par la suite on utilisera la formulation plus synthétique suivante, pour $t \in [0, 2\pi]$:

$$X^{(n)}(t) = \sum_{i=1}^n \rho_i(t)X_i, \quad (2.25)$$

En vérité, on peut voir que pour tout $t \in [0, 2\pi]$, les $\rho_i(t)$ sont les coordonnées sphériques d'un point de la sphère unité de dimension n . On a vu précédemment (section 2.3) comment représenter un processus du second ordre en utilisant la décomposition KL et comment l'approximer en tronquant la base de vecteurs propres à un niveau α donné. L'erreur commise est alors une combinaison linéaire des vecteurs restant. On a ainsi la représentation suivante :

$$X_u = \sum_{i=1}^{N_{1-\alpha}} \sqrt{\lambda_i} \xi_i \phi_i(u) + R(u), \quad (2.26)$$

où $R(u)$ représente l'erreur d'approximation et est tel que $\text{Var}(R(u)) \leq \alpha$. Si l'on injecte (2.26) dans (2.25), on obtient :

$$X_u^{(n)}(t) = \sum_{i=1}^n \rho_i(t) \left[\sum_{j=1}^{N_{1-\alpha}} \sqrt{\lambda_j} \xi_j^{(i)} \phi_j(u) + R_i(u) \right], \quad (2.27)$$

2.6. AUTRES APPLICATIONS DE LA DÉCOMPOSITION DE KARHUNEN-LOÈVE

En notant :

$$R^{(n)}(u) = \sum_{i=1}^n \rho_i(t) R_i(u),$$

on peut réécrire :

$$X_u^{(n)}(t) = \sum_{j=1}^{N_{1-\alpha}} \sqrt{\lambda_j} e_j(u) \sum_{i=1}^n \rho_i(t) \xi_j^{(i)} + R^{(n)}(u), \quad (2.28)$$

où par linéarité et indépendance on a $\text{Var}(R^{(n)}(u)) \leq \alpha$. Dans l'approximation du processus X par la décomposition tronquée de KL, on a vu que l'on conserve $N_{1-\alpha}$ variables ξ_i pour la description de X . Autrement dit, l'espace des variables aléatoires qu'il s'agira de caler est de dimension $N_{1-\alpha}$, on peut l'écrire de la manière suivante : $F^{-1}([0, 1]^{N_{1-\alpha}})$, où F est la fonction de répartition d'une variable gaussienne. Couvrir un sous-ensemble d'un tel espace via la paramétrisation par déformation graduelle nécessite $N_{1-\alpha}$ paramètres $t_1, \dots, t_{N_{1-\alpha}}$ et donc il nous faut une réalisation x_i de plus que le nombre de paramètres d'après l'expression générale (2.24). Cela revient à projeter une sphère de dimension $(N_{1-\alpha} + 1)$ sur une boule de dimension $N_{1-\alpha}$.

Ainsi, pour un α donné, si l'on choisit $n = N_{1-\alpha} + 1$, on couvrira un hypervolume de dimension n qui représente approximativement $1 - \alpha\%$ de la variance totale de l'échantillon (x_1, \dots, x_n) , dans un sens qu'il reste à définir.

Il est important de préciser ce point, à savoir de pouvoir calculer la mesure du volume couvert par une telle combinaison de réalisations. Elle dépend en effet fortement de la manière dont les x_1, x_2, \dots, x_n sont échantillonnés. Si l'on considère un cas simple avec un nombre de mailles très faible, disons 10. On fait l'hypothèse que seuls 2 paramètres sont significatifs au niveau de confiance 95% dans la décomposition KL. Ainsi, on peut couvrir 95% de la variance de l'échantillon en ne considérant que des combinaisons de 3 réalisations. Si ces 3 réalisations se trouvent sur une même droite dans un plan de dimension 2, la surface couverte par la combinaison $X_u^{(3)}(t)$ sera de mesure nulle. Ceci illustre parfaitement le problème de l'échantillonnage via la déformation graduelle. Une manière de pallier ce problème est d'employer des techniques de planification d'expérience, entre autre, est d'effectuer un échantillonnage en hypercube latin en maximisant la distance minimale entre les points. On peut également penser générer les réalisations à partir de suites à discrétion faible [Fang and Wang, 1994].

Enfin, il est important de noter la grande différence entre ces deux paramétrisations : la méthode de déformation graduelle est une paramétrisation de l'espace d'incertitude, dans le sens où elle négligera les réalisations qui ne sont pas de la forme (2.24), ne couvrira donc pas totalement le support de la loi a priori, et ne restituera pas la probabilité associée à chaque réalisation (en toute rigueur, cette probabilité est nulle, nous l'entendons ici dans le sens où, par exemple, une réalisation gaussienne plus proche de la moyenne qu'une autre se verra attribuée une probabilité plus forte). En revanche, la décomposition KL va négliger les variations de faible amplitude en variance mais conservera les probabilités associées aux réalisations. En définitive, on peut voir une réalisation tronquée dans la base de vecteur propre comme une réalisation d'un modèle géostatistique qui sera une version simplifiée du modèle géologique. C'est cette approche que nous développons dans la suite.

2.6.2 Karhunen-Loève comme méthode d'approximation

Apport du gradient

Supposons que l'on a une réalisation approchée calée \hat{X}_u^N à l'issue d'une procédure d'optimisation ou d'une approche par échantillonnage comme ci-dessus. Elle se présente sous la forme suivante :

$$\hat{X}_u^N = \sum_{i=1}^N \sqrt{\lambda_i} \xi_i \phi_i(u), \text{ où } u \in \{u_1, \dots, u_H\}, H \text{ nombre de mailles du champ.}$$

On sait que cette quantité approche un ensemble de réalisations du modèle complet (non-tronqué), de la manière suivante :

$$\begin{aligned} \hat{X}_u &= \hat{X}_u^{(N)} + R_u^{(N)}, \\ \text{où : } R_u^{(N)} &\sim \mathcal{N}(0, \Gamma - \Gamma^{(N)}), \\ \Gamma^{(N)} &= \Phi^{(N)} \Lambda \Phi^{(N)t}, \\ \Phi^{(N)} &\text{ est la matrice } L \times N \text{ dont les vecteurs colonnes sont les } \phi_i(x), \\ \Lambda &\text{ est la matrice diagonale } N \times N \text{ dont les éléments diagonaux sont les } \lambda_i. \end{aligned}$$

On notera $\Gamma_R^{(N)} = \Gamma - \Gamma^{(N)}$.

On peut alors associer une distribution à $F(\hat{X})$ par un développement de Taylor à l'ordre 1 :

$$F(\hat{X}) \approx F(\hat{X}_u^{(N)}) + R_u^{(N)} \nabla F(\hat{X}_u^{(N)}) \sim \mathcal{N}(\hat{X}_u^{(N)}, \nabla F(\hat{X}_u^{(N)})^t \Gamma_R^{(N)} \nabla F(\hat{X}_u^{(N)}))$$

Ainsi, à partir de chaque réalisation tronquée calée, connaissant le gradient de F en ce point, on peut appréhender un ensemble de réalisations complètes, connaissant la loi du reste $R_u^{(N)}$, et évaluer, sans effectuer de simulation d'écoulement, les résultats d'écoulement pour chaque élément cet ensemble.

Découpage du champ et simulation par blocs

En pratique, lorsque l'on cherche à résoudre un problème d'history matching, on peut se rendre compte que certaines zones seront moins influentes que d'autres sur les résultats de la simulation d'écoulement. On applique alors souvent le principe de "zonation" qui consiste à diviser le réservoir en plusieurs zones et à effectuer le calage sur chacune de ces zones, ce qui entraîne des discontinuités spatiales dans le modèle. Dans [Hu, 2000] est démontré que la méthode de déformation graduelle permet de déformer des zones restreintes du champ considéré tout en préservant la continuité spatiale. On va s'attacher à montrer ici que ce découpage du champ peut être effectué dans le cadre de la décomposition KL des champs aléatoires gaussiens. En outre, ce découpage du champ permet de réduire la taille des matrices dont on doit extraire les valeurs propres et vecteurs propres, rendant praticable l'application de cette paramétrisation pour de grands champs.

Il suffira en effet de découper la matrice de covariance en blocs correspondant chacun

2.6. AUTRES APPLICATIONS DE LA DÉCOMPOSITION DE KARHUNEN-LOÈVE

à une zone. Par exemple, considérons un champ aléatoire gaussien X de moyenne μ et de matrice de covariance Γ . On découpe ce champ en 2 blocs X_1 et X_2 . On peut alors écrire Γ et μ de la manière suivante :

$$\Gamma = \left(\begin{array}{c|c} \Gamma_1 & \Gamma_{1,2} \\ \hline \Gamma_{2,1} & \Gamma_2 \end{array} \right), \quad \mu^t = (\mu_1, \mu_2)^t.$$

On peut alors effectuer une simulation du champ comme suit :

1. on simule $X_1 \sim \mathcal{N}(\mu_1, \Gamma_1)$ et on obtient sa réalisation x_1
2. puis on simule $(X_2|X_1 = x_1) \sim \mathcal{N}(\mu_{(2|1)} = \mu_2 + \Gamma_{2,1}^t \Gamma_1^{-1} x_1, \Gamma_{2|1} = \Gamma_2 - \Gamma_{2,1} \Gamma_1^{-1} \Gamma_{1,2})$,

où $\mu_{(2|1)} = \mathbb{E}(X_2|X_1 = x_1)$ et $\Gamma_{2|1} = \text{var}(X_2|X_1 = x_1)$, par abus de notation. On peut généraliser la méthode à n blocs. La difficulté de cette méthode est de nature numérique et réside dans l'inversion de la matrice Γ_1 . Lorsque le nombre de blocs considérés est égal à n , Γ s'écrit :

$$\Gamma = \begin{pmatrix} \Gamma_1 & \Gamma_{1,2} & \cdots & \Gamma_{1,n} \\ \Gamma_{2,1} & \Gamma_2 & \cdots & \Gamma_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{n,1} & \Gamma_{n,2} & \cdots & \Gamma_n \end{pmatrix}$$

Si l'on veut simuler le troisième bloc X_3 , il faut calculer la matrice de covariance conditionnelle $\Gamma_{3|1,2}$ qui prend la forme suivante :

$$\Gamma_{3|1,2} = \Gamma_3 - \Gamma_{3,1:2} \Gamma_{1:2}^{-1} \Gamma_{1:2,3},$$

où $\Gamma_{1:2} = \begin{pmatrix} \Gamma_1 & \Gamma_{1,2} \\ \Gamma_{2,1} & \Gamma_2 \end{pmatrix}$, $\Gamma_{3,1:2} = (\Gamma_{3,1}, \Gamma_{3,2})$ et $\Gamma_{1:2,3} = \Gamma_{3,1:2}^t$. Il faut ainsi inverser la matrice $\Gamma_{1:2}$ de taille : (taille de X_1) \times (taille de X_2). On peut montrer que la dimension de ce problème peut se réduire à celle de 2 en employant la formule suivante d'inversion par bloc :

$$\begin{aligned} \Gamma_{1:2}^{-1} &= \begin{pmatrix} \Gamma_1 & \Gamma_{1,2} \\ \Gamma_{2,1} & \Gamma_2 \end{pmatrix}^{-1} \\ &= \begin{pmatrix} \Gamma_1^{-1} + \Gamma_1^{-1} \Gamma_{1,2} (\Gamma_{2|1})^{-1} \Gamma_{2,1} \Gamma_1^{-1} & -\Gamma_1^{-1} \Gamma_{1,2} (\Gamma_{2|1})^{-1} \\ -(\Gamma_{2|1})^{-1} \Gamma_{2,1} \Gamma_1^{-1} & (\Gamma_{2|1})^{-1} \end{pmatrix} \end{aligned}$$

Ainsi, il suffit d'inverser la matrice $\Gamma_{2|1}$ de la taille de 2. La moyenne conditionnelle de $X_3|X_1 = x_1, X_2 = x_2$ s'écrit :

$$\mu_{i|1,2} = \mu_3 + \Gamma_{3,1:2}^t \Gamma_{1:2}^{-1} x_{1:2},$$

où $x_{1:2} = (x_1, x_2)^t$.

A l'itération $i + 1$, on aura déjà calculé et stocké $\Gamma_{1:(i-1)}^{-1}$ et pour calculer :

$$\Gamma_{i+1|1,\dots,i} = \Gamma_{i+1} - \Gamma_{i+1,1:i} \Gamma_{1:i}^{-1} \Gamma_{1:i,i+1},$$

le problème d'inversion portera sur $\Gamma_{1:i}$. On le résoudra en appliquant la formule d'inversion par blocs :

$$\begin{aligned}\Gamma_{1:i}^{-1} &= \begin{pmatrix} \Gamma_{1:(i-1)} & \Gamma_{1:(i-1),i} \\ \Gamma_{1:(i-1),1} & \Gamma_i \end{pmatrix}^{-1} = \begin{pmatrix} A & B \\ B^t & C \end{pmatrix}, \\ \text{où} \quad A &= \Gamma_{1:(i-1)}^{-1} + \Gamma_{1:(i-1)}^{-1} \Gamma_{1:(i-1),i} (\Gamma_{i|1,\dots,i-1})^{-1} \Gamma_{i,1:(i-1)} \Gamma_{1:(i-1)}^{-1} \\ B &= -\Gamma_{1:(i-1)}^{-1} \Gamma_{1:(i-1),i} (\Gamma_{i|1,\dots,i-1})^{-1} \\ C &= \Gamma_{i|1,\dots,i-1})^{-1}\end{aligned}$$

Ainsi, il suffira cette fois d'inverser la matrice $\Gamma_{i|1,\dots,i-1}$ de la taille du bloc i , puisque la matrice $\Gamma_{1:(i-1)}^{-1}$ est connue. La moyenne conditionnelle de $X_i + 1 | X_1 = x_1, \dots, X_i = x_i$ prendra la forme suivante :

$$\mu_{i|1,\dots,i-1} = \mu_i + \Gamma_{i,1:(i-1)}^t \Gamma_{1:(i-1)}^{-1} x_{1:(i-1)},$$

où $x_{1:(i-1)} = (x_1, \dots, x_{i-1})^t$.

On peut ensuite effectuer la décomposition KL sur chaque bloc. A chaque itération, on peut remarquer que l'algorithme nécessite une inversion de matrice et quatre produits matriciels. Une étude de sa stabilité est à envisager.

Cette méthode présente l'avantage que les blocs peuvent être choisis de tailles diverses et que le nombre de composantes retenu peut varier d'un bloc à l'autre. On peut ainsi garder, par exemple, un nombre plus important de composantes aux abords des puits, où les faibles hétérogénéités auront théoriquement plus d'impact, que dans les zones du réservoir de moindre intérêt. A noter également qu'en théorie, le nombre de composantes *significatives* pour chaque bloc sera plus faible que si l'on effectue la décomposition sur l'ensemble du champ. En effet, le rapport portée pratique sur longueur maximum au sein du bloc sera augmenté et les résultats de la partie 2.4.1 montrent que le nombre de composantes nécessaires à reproduire un pourcentage donné de la variance s'en trouvera diminué.

2.6.3 Approches alternatives

Approche empirique, vers les modèles de géostatistique multipoints

Une autre manière de procéder, et qui ne se limite pas aux modèles gaussiens, peut consister à générer une ou plusieurs réalisations du modèle considéré (multipoint, objet,...) puis calculer la décomposition KL de la covariance empirique de l'échantillon, à l'image des travaux de [Kirby and Sirovich, 1995]. Une telle paramétrisation a déjà été proposée dans [Sarma et al., 2006], avec une approche légèrement différente. Dans cet article, les auteurs associent un modèle de covariance à l'image d'apprentissage à partir de laquelle les réalisations en multipoint sont générées à la suite de l'étape de reconnaissance de forme, enfin ce modèle de covariance est décomposé selon la méthode KL et les paramètres résultants sont calés par rapport aux données de production. Le problème dans cette méthodologie vient de l'approximation de l'image multipoint par un processus gaussien, il nous semble que l'on perd à terme l'aspect multipoint du modèle. En effet, le cadre gaussien est idéal pour l'emploi de ce type de décomposition

2.6. AUTRES APPLICATIONS DE LA DÉCOMPOSITION DE KARHUNEN-LOÈVE

dans le sens où l'on sait qu'un processus gaussien est entièrement déterminé par ses moments d'ordre 1 et 2. Dans le cas contraire, il convient d'inférer les distributions des composantes dans la décomposition de KL ; on a vu au paragraphe 2.5.3 la difficulté d'une telle entreprise, qui interdit l'emploi d'une méthode d'échantillonnage du type de celles développées dans le chapitre 3.

En outre, les modèles multipoints [Strebelle, 2002] sont déterminés par les corrélations d'ordre supérieur, dans le sens où la connaissance des deux premiers moments ne permet pas de déterminer entièrement le modèle. Il pourrait dès lors être intéressant d'étendre ce principe de décomposition aux moments d'ordres supérieurs, à l'image de la méthode de l'*Independent Component Analysis* [Hyvärinen and Oja, 2000], méthode déjà appliquée avec succès sur le problème de la caractérisation des visages [Alex et al., 2002]. Ces méthodes issues de l'algèbre multilinéaire [De Lathauwer, 1997] posent un cadre théorique qui est parfaitement adapté à l'extension du principe de la décomposition KL aux moments d'ordre supérieurs et pourraient permettre de s'affranchir du cadre gaussien. Le coût informatique de ces méthodes semble néanmoins prohibitif : il ne s'agit plus de diagonaliser une matrice de covariance mais des tenseurs de moments de dimension 3, 4 voire au-delà. Compte-tenu du nombre de mailles des modèles de réservoir, cette approche semble déraisonnable. Dans cette ordre d'idée, on peut également citer la méthode dite *kernel PCA* (analyse en composantes principales à noyaux) proposée dans [Sarma et al., 2007]. Mais, encore une fois, la loi des composantes de la réalisation dans la nouvelle base est inconnue et une méthode d'échantillonnage n'est pas applicable.

EDP stochastiques

Dans ce cadre, l'approche est radicalement différente : le simulateur d'écoulement n'est plus considéré comme une boîte noire sur laquelle on va réaliser une analyse d'incertitude. Au contraire, on va placer l'aléa directement au coeur des équations d'écoulement, que l'on va projeter sur une base tronquée de vecteurs propres issue de la décomposition KL, voire sur les fonctions propres des moments d'ordre supérieur. Il est particulièrement délicat de résoudre les équations d'écoulement dans cette nouvelle base puisque cela revient à multiplier le nombre d'équations du modèle d'écoulement par le nombre de composantes retenues dans la décomposition de KL. Cependant, on peut tirer partie de cette formulation comme proposé par [Zhang and Lu, 2004] pour un modèle d'écoulement monophasique puis [Chen et al., 2005] pour un modèle bi-phasique. L'idée est de représenter les grandeurs de pression et de saturation sous la forme de polynômes de chaos (voir [Ghanem and Spanos, 1991]), fonctions des composantes ξ_i employées dans la décomposition KL de la perméabilité. Un certain nombre de simulations d'écoulement permettent alors d'estimer les paramètres des polynômes de chaos, ce qui fournit ainsi un émulateur du simulateur d'écoulement : pour des valeurs données des composantes ξ_i , on pourra calculer à la fois le champ de perméabilité (approximé) et les valeurs de pression et de saturation correspondant. Néanmoins, cette approche nécessite de fixer arbitrairement un certain nombre de paramètres, parmi lesquels le nombre de composantes employées dans la décomposition KL et l'ordre des polynômes de chaos utilisés pour représenter la pression et la saturation, un nombre trop faible risquant d'entraîner des résultats peu réalistes.

Chapitre 3

Méthodes de Monte-Carlo par chaînes de Markov : application au problème inverse

3.1 Introduction

Comme on l'a vu dans l'introduction du manuscrit, le problème du calage d'historique en ingénierie de réservoir est un problème inverse mal posé. Sa résolution implique d'identifier les réalisations du modèle géologique X de réservoir correspondant aux données de production obtenues lors de la phase de production du champ. Autrement dit, il s'agit d'inférer la loi $\mathbb{P}(X|d)$ du modèle géologique conditionnée aux données de production. Le modèle géologique (porosité, perméabilité, faciès) est la plupart du temps obtenu à partir d'un modèle de champ gaussien, dont les moments d'ordre 1, μ , et 2, Γ , sont supposés connus. Les données de production d sont en général entachées d'une erreur englobant l'erreur de mesure et l'erreur commise par le simulateur d'écoulement F , l'opérateur direct. Cette erreur est également supposée gaussienne, de moyenne nulle et de covariance C_d supposée connue. En général, cette matrice de covariance est diagonale et ses éléments associent des poids différents aux données disponibles. La formulation de la loi conditionnelle est alors donnée via le théorème de Bayes de la manière suivante :

$$\mathbb{P}(X|d) \propto e^{\left(-\frac{1}{2}\|d-F(X)\|_{C_d^{-1}}^2 - \frac{1}{2}\|X-\mu\|_{\Gamma^{-1}}^2\right)}, \quad (3.1)$$

Les méthodes les plus efficaces pour échantillonner ce type de loi, malgré leur coût informatique sont les méthodes de Monte-Carlo par chaîne de Markov (McMC). En effet, les algorithmes McMC, introduits par [Metropolis et al., 1953] et largement décrits dans [Robert and Casella, 2004], sont des méthodes populaires pour générer des échantillons d'une distribution π sur un espace $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, théoriquement sans limitation sur π . La distribution cible décrit souvent la loi a posteriori dans un problème d'inférence bayésienne, comme ici (3.1). Typiquement, \mathcal{X} est de très grande dimension, en général, $\mathcal{X} \subseteq \mathbb{R}^d$, où d est grand, et/ou π a une expression si complexe que son calcul direct est impossible ; la constante de normalisation est en général incalculable, en particulier dans la problématique de l'inversion bayésienne. Face à un tel problème, une méthode McMC consiste à générer une chaîne de Markov ergodique $\{X_k, k \geq 0\}$ sur

\mathcal{X} , de noyau de transition K dont la loi stationnaire est π , c'est-à-dire pour $A \in \mathcal{B}(\mathcal{X})$:

$$\int_{\mathcal{X}} K(x, A) \pi(x) dx = \pi(A) \quad (3.2)$$

Cependant, dans le cas où la loi à inférer a une structure complexe, *e.g.* est fortement multimodale, ces méthodes n'effectuent qu'une description locale de (3.1) dans un temps raisonnable. En particulier, elles présentent des difficultés à identifier les éventuels différents modes du posterior, autrement dit à couvrir de manière satisfaisante le support de la loi a posteriori. A la lumière d'un exemple très simple de problème inverse, nous illustrons les problèmes posés par l'emploi de différents algorithmes MCMC dans le cadre de la résolution de problèmes inverse en section 3.5.

Il convient alors d'adopter des stratégies permettant à la chaîne de se déplacer plus facilement sur le support de (3.1) et d'ainsi mieux couvrir l'espace d'incertitude. Une solution, inspirée de la technique du recuit-simulé, passe par l'emploi d'un certain nombre de chaînes en parallèles, chacune à une température différente, entre lesquelles des interactions auront lieu. Nous en décrivons le principe dans la section 3.6 et une application sur le problème inverse précédent dans le paragraphe 3.6.6.

Après avoir introduit les principales motivations de l'emploi des méthodes de Monte-Carlo (section 3.2), nous présentons les notions fondamentales des chaînes de Markov et les résultats nécessaires à établir la convergence de nombreux algorithmes MCMC dans la section 3.3. Ensuite, nous nous attacherons à décrire l'algorithme de Metropolis-Hastings et ses propriétés (section 3.4). Ces deux parties sont fortement inspirées des chapitres 6 et 7 de [Robert and Casella, 2004]. En outre, nous présentons deux méthodes plus récentes : les chaînes adaptatives, dans le paragraphe 3.4.7 et les chaînes interagissantes en section 3.6.

3.2 Motivations : calcul d'intégrales et de sommes

Les méthodes de simulation sont basées sur la production de nombres aléatoires, distribués selon une loi de probabilité π dont la forme explicite n'est pas nécessairement connue. Dans cette section, on suppose que toutes les lois de probabilité admettent une densité par rapport à la mesure de Lebesgue et que $\mathcal{X} \subset \mathbb{R}^d$.

Dans de nombreuses applications, pour une certaine fonction h , on souhaite calculer

$$\mathcal{I} = \mathbb{E}_{\pi} [h(X)] = \int h(x) \pi(x) dx, \quad (3.3)$$

Dans le domaine pétrolier, $\pi(x)$ représente par exemple la densité de probabilité spatiale de la perméabilité du réservoir et h une prédiction de la production d'huile.

En général, même si l'on sait évaluer h en tout point, on ne peut calculer formellement l'intégrale \mathcal{I} . Le calcul d'intégrale par la méthode Monte-Carlo consiste dans sa version la plus simple à générer $(X_1, \dots, X_n) \sim_{i.i.d.} \pi$, et à approcher (3.3) par la moyenne empirique $S_n(h) = \frac{1}{n} \sum_{i=1}^n h(x_i)$, où i.i.d signifie indépendant et identiquement distribué. En effet, d'après la loi forte des grands nombres, si l'espérance de X est finie, on a

l'assurance que

$$S_n(h) \rightarrow_{p.s.} \int h(x)\pi(x)dx, \quad (3.4)$$

où p.s. signifie presque sûrement. Si de plus, on sait que si

$$\int h(x)^2\pi(x)dx < \infty, \quad (3.5)$$

la vitesse de convergence de $S_n(h)$ peut être évaluée, puisque la variance

$$\text{var}(S_n(h)) = \frac{1}{n} \int (h(x) - \mathbb{E}_\pi[h(X)])^2 \pi(x)dx \quad (3.6)$$

peut également être estimée à partir de l'échantillon (X_1, \dots, X_n) par la quantité

$$\sigma_n^2 = \frac{1}{n} \sum_{i=1}^n (h(x_i) - S_n(h))^2. \quad (3.7)$$

Le théorème de la limite centrale nous assure alors que pour n grand,

$$\frac{(S_n(h) - \mathbb{E}_\pi[h(X)])}{\sigma_n} \quad (3.8)$$

suit approximativement une loi $\mathcal{N}(0, 1)$. Cette propriété conduit à la construction de tests de convergence et de bornes de confiance asymptotiques pour $S_n(h)$. En outre, elle nous indique que la vitesse de convergence de $S_n(h)$ est de l'ordre de \sqrt{n} et ce indépendamment de la dimension du problème. Cela explique l'efficacité de cette méthode par rapport aux méthodes d'intégration numérique déterministes dont les vitesses de convergence décroissent rapidement avec la dimension du problème.

Les méthodes classiques de génération de variables aléatoires sont largement répandues et bien décrites dans le chapitre 2 de [Robert and Casella, 2004], avec un grand nombre d'exemples. Pour autant, l'application de cette méthode repose sur notre capacité à générer un échantillon $(X_1, \dots, X_n) \sim_{i.i.d.} \pi$ et dans la plupart des problèmes inverses, la complexité de la loi (3.1) (dimension, support) est telle que des méthodes plus avancées sont nécessaires.

On peut aussi citer les méthodes d'échantillonnage d'importance, développées dans le chapitre 3 de [Robert and Casella, 2004] qui reposent sur l'emploi d'une fonction d'importance g . Son principe est décrit dans la définition suivante :

Définition 3.2.1 *La méthode d'échantillonnage d'importance est une évaluation de (3.3) basée sur la simulation d'un échantillon X_1, \dots, X_n de loi g et approximant :*

$$\mathbb{E}_\pi[h(X)] \approx \frac{1}{n} \sum_{i=1}^n \frac{\pi(X_i)}{g(X_i)} h(X_i) \quad (3.9)$$

Cette méthode est basée sur la représentation suivante de (3.3) :

$$\mathbb{E}_\pi[h(X)] = \int h(x) \frac{\pi(x)}{g(x)} g(x)dx, \quad (3.10)$$

On peut montrer (cf. [Robert and Casella, 2004]) qu'il existe un choix de g qui minimise la variance de (3.9). Néanmoins, cette méthode, en principe universelle, repose sur le choix de la fonction d'importance, dont la difficulté s'accroît en même temps que la dimension du problème. D'autre part, le choix optimal de g dépend de la fonction h . Dans le cadre de l'inversion bayésienne, la ou les fonctions h d'intérêt sont généralement liées à l'opérateur direct, ce qui complique encore le choix de g . Pour autant, cette notion nous sera utile dans la section 3.6.

Nous reprenons donc le commentaire précédent, statuant la nécessité d'employer des méthodes de simulations plus élaborées pour résoudre les problèmes d'inversion bayésienne en général, et le problème du calage d'historique en particulier. En effet, les méthodes présentées dans cette section sont utiles uniquement lorsque l'on sait simuler directement selon la loi cible, où une loi l'approchant dans le cas de l'échantillonnage d'importance. Dans la section suivante, nous rappelons certains résultats classiques sur les chaînes de Markov qui nous permettront ensuite d'introduire l'algorithme de Metropolis-Hastings.

3.3 Chaînes de Markov

Dans cette section, nous introduisons les notions fondamentales des chaînes de Markov et les résultats nécessaires à établir la convergence de nombreux algorithmes MCMC. Nous ne proposons donc pas un exposé exhaustif sur les chaînes de Markov. En particulier, nous n'abordons pas le problème des chaînes de Markov à temps continu (ou processus de Markov) puisque le principe de la simulation nous conduit à considérer uniquement des processus stochastiques à temps discret¹.

3.3.1 Notions de base

Dans le cadre des algorithmes MCMC, les chaînes de Markov sont construites à partir d'un *noyau de transition* K (Définition 3.3.1), une densité de probabilité conditionnelle telle que $X_{n+1} \sim K(X_n, X_{n+1})$, où $X_n \in \mathcal{X}$ et \mathcal{X} est un espace topologique abstrait muni de sa tribu borélienne $\mathcal{B}(\mathcal{X})$.

Une chaîne de Markov est une suite de variables aléatoires qui évoluent au cours du temps (au cours des itérations) selon une certaine probabilité de transition. Il est donc naturel de définir une chaîne à partir de la notion de *noyau de transition*.

Définition 3.3.1 *Un noyau de transition est une fonction K définie sur $\mathcal{X} \times \mathcal{B}(\mathcal{X})$ telle que*

- i) $\forall x \in \mathcal{X}, K(x, \cdot)$ est une densité,
- ii) $\forall A \in \mathcal{B}(\mathcal{X}), K(\cdot, A)$ est mesurable.

¹Nous verrons plus loin avec (3.33) que certains algorithmes de Monte-Carlo emploient une représentation sous forme de diffusion pour accélérer la convergence vers la loi stationnaire

3.3. CHAÎNES DE MARKOV

Quand \mathcal{X} est un espace discret, le noyau de transition est simplement une matrice de transition d'éléments

$$K(x, y) = K_{xy} = \mathbb{P}(X_n = y | X_{n-1} = x), \quad \forall x, y \in \mathcal{X}.$$

Dans le cas continu, quand $\mathcal{X} \subset \mathbb{R}^d$, le noyau définit aussi la densité conditionnelle $K(x, y)$ par rapport à la mesure de Lebesgue de la transition $K(x, \cdot)$, c'est-à-dire $\mathbb{P}(Y \in A | x) = \int_A K(x, y) dy$.

La chaîne (X_n) est généralement indexée sur \mathbb{N} plutôt que sur \mathbb{Z} . Par conséquent, la distribution de X_0 , que l'on notera μ , l'état initial de la chaîne, joue un rôle important. On notera respectivement \mathbb{P}_x et \mathbb{P}_μ , la loi de probabilité sachant $X_0 = x$ et sachant $X_0 \sim \mu$. De même, \mathbb{E}_x et \mathbb{E}_μ désignerons l'espérance conditionnelle sachant $X_0 = x$, et sachant $X_0 \sim \mu$.

Définition 3.3.2 *Etant donné un noyau de transition K_n , une suite X_0, \dots, X_n, \dots de variables aléatoires est une chaîne de Markov, notée (X_n) , si pour tout n , la distribution conditionnelle de X_n sachant $X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \dots, X_0 = x_0$ est égale à la distribution conditionnelle de X_n sachant x_{n-1} , c'est-à-dire*

$$\begin{aligned} \mathbb{P}(X_n \in A | x_{n-1}, x_{n-2}, \dots, x_0) &= \mathbb{P}(X_n \in A | x_{n-1}) \\ &= \int_A K_n(x_n, dx) \end{aligned} \quad (3.11)$$

La chaîne est dite homogène si $K_n = K$, autrement dit si K ne varie pas au cours du temps.

Ainsi, dans le cas homogène, si la distribution initiale μ ou l'état initial $X_0 = x_0$ est connu, la loi de la chaîne de Markov ne dépend que de ses transitions : la distribution de X_n sachant $X_{n-1} = x_{n-1}$. Dans la suite, sauf mention contraire, nous considérons que (X_n) est une chaîne de Markov homogène. Le processus de marche aléatoire, défini comme suit, en est un bon exemple.

Définition 3.3.3 *Une suite de variables aléatoires (X_n) est une marche aléatoire si elle satisfait*

$$X_{n+1} = X_n + \varepsilon_n,$$

où ε_n est une suite de variables aléatoire i.i.d. indépendantes de X_n, X_{n-1}, \dots . Si la distribution de ε_n est symétrique par rapport à zéro, la suite est une marche aléatoire symétrique, ou à accroissements symétriques.

Il existe de nombreux exemples de marche aléatoire ; elles jouent un rôle clef dans de nombreux algorithmes McMC, notamment ceux inspirés de l'algorithme de Metropolis-Hastings (voir section 3.4). Les chaînes employées dans le cadre des algorithmes McMC possèdent une très forte propriété de stabilité ; l'existence d'une loi de probabilité stationnaire est assurée par construction.

Nous introduisons ici la notion de *temps d'arrivée* dans un sous-ensemble $A \in \mathcal{B}(\mathcal{X})$, essentielle dans l'analyse de la convergence des algorithmes McMC.

Définition 3.3.4 Soit $A \in \mathcal{B}(\mathcal{X})$. Le premier n où la chaîne entre dans l'ensemble A , $\tau_A = \inf\{n \geq 1; X_n \in A\}$, est le temps d'arrivée en A . Par convention, $\tau_A = \infty$ si $X_n \notin A$ pour tout n .

On peut également définir

$$\eta_A = \sum_{n=1}^{\infty} 1_A(X_n), \quad (3.12)$$

le nombre de passage de (X_n) en A .

Les quantités $\mathbb{E}_x[\eta_A]$ et $\mathbb{P}_x(\tau_A < \infty)$, respectivement le nombre moyen de passages en A et la probabilité de retour en A , sont d'une importance particulière.

3.3.2 Propriétés

Les chaînes que l'on rencontre dans le cadre des algorithmes MCMC possèdent des propriétés spécifiques qui permettent d'établir les principaux résultats de convergence présentés ici.

Irréductibilité

La propriété d'irréductibilité est une première mesure de la sensibilité de la chaîne de Markov (X_n) aux conditions initiales : $X_0 = x_0$ ou $X_0 \sim \mu$. Elle est cruciale dans le cadre des algorithmes MCMC dans la mesure où elle induit une garantie de convergence.

Définition 3.3.5 Etant donnée une mesure ϕ , une chaîne de Markov (X_n) est dite ϕ -irréductible, si $\forall A \in \mathcal{B}(\mathcal{X})$, il existe $n \in \mathbb{N}$, tel que $\phi(A) > 0$ et $\forall x \in \mathcal{X}$:

$$\mathbb{P}_x(X_n \in A) > 0, \text{ autrement dit } K^n(x, A) > 0, \quad (3.13)$$

où $K^n(x, \cdot)$ dénote le noyau pour n transitions. La chaîne est fortement ϕ -irréductible, si $n=1$ pour tout A .

Atomes et ensembles petits

Alors que dans le cas discret, le noyau de transition est nécessairement atomique dans le sens usuel, c'est-à-dire qu'il existe des points de l'espace d'état de masse positive, la généralisation de cette notion au cas continu nécessite l'introduction de deux nouvelles notions :

Définition 3.3.6 La chaîne de Markov (X_n) a un atome $\alpha \in \mathcal{B}(\mathcal{X})$ s'il existe une mesure non-nulle ν associée telle que

$$K(x, A) = \nu(A), \quad \forall x \in \alpha, \forall A \in \mathcal{B}(\mathcal{X}).$$

Si (X_n) est ϕ -irréductible, l'atome est accessible quand $\phi(A) > 0$.

3.3. CHAÎNES DE MARKOV

Bien que triviale dans le cas discret, cette notion est souvent trop forte pour être employée dans le cas continu puisqu'elle implique que le noyau de transition soit constant sur un ensemble de mesure positive. Une généralisation plus efficace est donnée par la *condition minorante* : il existe un ensemble $C \in \mathcal{B}(\mathcal{X})$, une constante $\varepsilon > 0$ et une mesure de probabilité ν tels que

$$K(x, A) \geq \varepsilon \nu(A), \quad \forall x \in C, \quad \forall A \in \mathcal{B}(\mathcal{X}) \quad (3.14)$$

La mesure de probabilité ν apparaît alors comme une composante constante du noyau de transition sur C . (3.14) conduit à la notion d'*ensembles petits* (*small sets*) et au théorème qui suit :

Définition 3.3.7 *Un ensemble C est dit petit s'il existe $m \in \mathbb{N}^*$ et une mesure non nulle ν_m tels que*

$$K^m(x, A) \geq \nu_m(A), \quad \forall x \in C, \quad \forall A \in \mathcal{B}(\mathcal{X}). \quad (3.15)$$

Le résultat suivant indique la connection entre ensembles petits et irréductibilité.

Théorème 3.3.1 *Soit (X_n) une chaîne de Markov ϕ -irréductible. Pour tout ensemble $A \in \mathcal{B}(\mathcal{X})$, tel que $\phi(A) > 0$, il existe $m \in \mathbb{N}^*$ et un ensemble petit $C \subset A$ tels que la condition minorante associée satisfasse $\nu_m(C) > 0$. De plus \mathcal{X} peut être décomposé en une partition dénombrable d'ensembles petits.*

Apériodicité

Le comportement de (X_n) peut parfois être restreint par des contraintes déterministes sur les mouvements de X_n à X_{n+1} . Ces contraintes se formalisent avec les notions de cycles et d'apériodicité.

Définition 3.3.8 *Une chaîne (X_n) irréductible a un cycle de longueur d s'il existe un ensemble petit C , un entier associé M et une distribution de probabilité ν_M tels que d est le p.g.c.d. de*

$$\{m \geq 1; \exists \delta_m \text{ tel que } C \text{ est petit pour } \nu_m \geq \delta_m \nu_M\}$$

La notion importante au-delà de cette définition est la propriété d'apériodicité, c'est-à-dire quand $d = 1$. Cette propriété est fondamentale pour la démonstration de la plupart des résultats de convergence.

Récurrence

D'un point de vue algorithmique, une chaîne de Markov doit posséder de bonnes propriétés de "stabilité" pour garantir une approximation satisfaisante du modèle simulé. Effectivement, l'irréductibilité assure que chaque ensemble A sera visité par la chaîne (X_n) avec une probabilité non-nulle, mais cette propriété est trop faible pour assurer que la trajectoire de (X_n) entrera en A assez souvent. Pour formaliser cette stabilité de la chaîne de Markov, on doit introduire la notion de récurrence.

Définition 3.3.9 Un ensemble A est dit récurrent si $\mathbb{E}_x[\eta_A] = \infty$ pour tout $x \in A$. L'ensemble A est uniformément transient s'il existe M tel que $\mathbb{E}_x[\eta_A] < M$ pour tout $x \in A$.

Définition 3.3.10 Une chaîne de Markov (X_n) est récurrente si

- i) Il existe une mesure ϕ telle que (X_n) est ϕ -irréductible, et
- ii) $\forall A \in \mathcal{B}(\mathcal{X})$ tel que $\phi(A) > 0$, $\mathbb{E}_x[\eta_A] = \infty \forall x \in A$.

La chaîne est *transiente* si elle est ϕ -irréductible et \mathcal{X} est transient. On peut montrer qu'une chaîne ϕ -irréductible est soit récurrente, soit transiente. Il y a ainsi une stricte dichotomie entre la transience et la récurrence des chaînes de Markov irréductibles. Le résultat suivant indique un critère pour déterminer la récurrence de (X_n) . Construit par analogie au cas discret, il est basé sur la notion d'ensembles petits.

Proposition 3.3.1 Une chaîne ϕ -irréductible (X_n) est récurrente, s'il existe un ensemble petit C , avec $\phi(C) > 0$ tel que $\mathbb{P}_x(\tau_C < \infty) = 1$ pour tout $x \in C$.

Une seconde méthode pour vérifier la récurrence est basée sur la généralisation de la notion d'ensemble petit et de condition minorante. Elle nécessite l'emploi d'une *fonction potentiel* V et d'une *condition de dérive*. Ces notions seront utiles dans le paragraphe 3.4.7 et plus loin, en section 3.6.

Une notion plus forte de récurrence, qui porte le nom de son auteur, implique non seulement un nombre de visites moyen infini de chaque ensemble petit mais également un nombre infini de visites de chaque trajectoire de la chaîne de Markov.

Définition 3.3.11 Une chaîne de Markov (X_n) est récurrente de Harris si

- i) il existe une mesure ϕ telle que (X_n) est ϕ -irréductible, et
- ii) $\forall A \in \mathcal{B}(\mathcal{X})$ tel que $\phi(A) > 0$, $\mathbb{P}_x(\eta_A = \infty) = 1 \forall x \in A$.

La récurrence de Harris assure que la chaîne a le même comportement limite pour tout x . Ce dernier point est très important dans le contexte des algorithmes MCMC, puisque la plupart des algorithmes est initialisée en un point x_0 arbitraire. En pratique, la plupart des algorithmes MCMC employés possèdent la propriété de récurrence de Harris.

3.3.3 Mesure invariante

Nous introduisons maintenant la notion de mesure invariante, qui justifie l'emploi des méthodes MCMC.

Définition 3.3.12 Une mesure π est invariante pour le noyau de transition $K(\cdot, \cdot)$, et pour la chaîne associée, si

$$\int_{\mathcal{X}} K(x, A) \pi(x) dx = \pi(A), \quad \forall A \in \mathcal{B}(\mathcal{X})$$

Quand il existe une mesure de probabilité invariante pour une chaîne ϕ -irréductible, la chaîne est dite *positive*. La distribution invariante est également appelée distribution *stationnaire* : si π est une mesure de probabilité, $X_0 \sim \pi$ implique que $X_n \sim \pi$ pour tout n et ainsi la chaîne est stationnaire.

On montre qu'une chaîne positive admet une mesure de probabilité invariante unique à une constante multiplicative près.

Proposition 3.3.2 *Si la chaîne (X_n) est positive, alors elle est récurrente.*

On parlera alors de chaînes *positives* et de chaînes *Harris positives*. En outre,

Proposition 3.3.3 *Si (X_n) est récurrente, il existe une mesure invariante qui est unique à une constante multiplicative près.*

Réversibilité

Enfin, la propriété de stabilité inhérente aux chaînes stationnaires peut être reliée à la propriété de stabilité appelée *réversibilité*. Cette dernière implique que la direction du temps n'influe pas sur la dynamique de la chaîne.

Définition 3.3.13 *Une chaîne de Markov (X_n) stationnaire est réversible si la distribution de X_{n+1} conditionnellement à $X_{n+2} = x$ est égale à la distribution de X_{n+1} conditionnellement à $X_n = x$.*

En fait, la réversibilité peut être liée à l'existence d'une mesure π satisfaisant une condition plus forte que la définition 3.3.12 :

Définition 3.3.14 *Une chaîne de Markov avec un noyau de transition K , admettant une densité positive $K(x, y)$ (par rapport à une mesure $\nu_{\mathcal{X}}$ sur \mathcal{X} , notée $K(x, dy) = K(x, y)\nu_{\mathcal{X}}(dy)$) satisfait la condition de balance détaillée s'il existe une fonction f telle que*

$$K(x, y)f(x) = K(y, x)f(y) \quad (3.16)$$

pour tout (x, y) .

Bien que cette condition ne soit pas nécessaire pour que f soit la densité par rapport à la mesure de Lebesgue sur \mathcal{X} d'une mesure stationnaire associée à K , elle est néanmoins suffisante et est souvent simple à vérifier pour la plupart des algorithmes MCMC. La condition de balance détaillée (3.16) exprime un équilibre dans le flux de la chaîne de Markov ; la probabilité d'être en x et d'aller en y est égale à la probabilité d'être en y et d'aller en x . Quand f est une densité, (3.16) implique aussi que la chaîne est réversible. Plus généralement,

Théorème 3.3.2 *Si une chaîne de Markov de noyau de transition K satisfait la condition de balance détaillée avec une densité de probabilité π , alors :*

- i) la densité π est la densité associée à la mesure invariante de la chaîne,*
- ii) la chaîne est réversible.*

3.3.4 Convergence et théorèmes limites

La distribution stationnaire est aussi une distribution limite dans le sens où la distribution limite de X_{n+1} est π sous la norme de variation totale, quelle que soit la valeur de x_0 . Cette norme est définie sur l'ensemble des mesures (finies) signées par :

$$\|\mu_1 - \mu_2\|_{TV} = \sup_A |\mu_1(A) - \mu_2(A)|.$$

Différents théorèmes de convergence existent. Nous retiendrons en priorité le suivant :

Théorème 3.3.3 *Si (X_n) est Harris positive et apériodique, alors*

$$\lim_{n \rightarrow \infty} \left\| \int K^n(x, \cdot) \mu(dx) - \pi \right\|_{TV} = 0$$

pour toute distribution initiale μ .

Convergence géométrique et uniforme

Des formes plus fortes de convergence peuvent intervenir dans le cadre des MCMC, comme les convergences géométrique et uniforme.

La notion de convergence géométrique nécessite l'introduction d'une extension de la norme de variation totale, notée $\|\cdot\|_h$. Elle est définie par :

$$\|\mu\|_h = \sup_{|g| < h} \left| \int_{\mathcal{X}} g(x) \mu(dx) \right|.$$

Définition 3.3.15 *Une chaîne (X_n) est géométriquement h -ergodique, avec $h \geq 1$ sur \mathcal{X} , si (X_n) est Harris positive, de distribution stationnaire π , si $\mathbb{E}_{\pi}[h] < \infty$, et s'il existe $r_h > 1$ tel que*

$$\sum_{i=1}^n r_h^i \|K^i(x, \cdot) - \pi\|_h < \infty$$

pour tout $x \in \mathcal{X}$. Quand $h = 1$ on dit que (X_n) est géométriquement ergodique.

La propriété de convergence uniforme est plus forte que l'ergodicité géométrique dans le sens où le taux de convergence géométrique doit être uniforme sur l'espace entier.

Définition 3.3.16 *La chaîne (X_n) est uniformément ergodique si*

$$\lim_{n \rightarrow \infty} \sup_{x \in \mathcal{X}} \|K^n(x, \cdot) - \pi\|_{TV} = 0.$$

L'ergodicité uniforme peut être établie par l'une des propriétés équivalentes suivantes :

Théorème 3.3.4 *Les conditions suivantes sont équivalentes :*

- i) (X_n) est uniformément ergodique ;*

ii) il existe $R < \infty$ et $r < 1$ tel que

$$\|K^n(x, \cdot) - \pi\|_{TV} < Rr^{-n}, \quad \forall x \in \mathcal{X};$$

iii) (X_n) est apériodique et \mathcal{X} est un ensemble petit;

iv) (X_n) est apériodique et il existe un ensemble petit C et un réel $\kappa > 1$ tel que

$$\sup_{x \in \mathcal{X}} \mathbb{E}_x[\kappa^{\tau_C}] < \infty,$$

où τ_C est le temps d'arrivée en C .

Théorèmes limites

Bien que les notions et résultats introduits dans les paragraphes précédents soient essentiels pour justifier l'emploi des algorithmes MCMC, le présent paragraphe est fondamental pour leur utilisation pratique. En effet, les différents résultats de convergence obtenus précédemment (ergodicité) déterminent principalement les propriétés probabilistes du comportement de la chaîne à l'instant n , lorsque n tend vers l'infini. Ces propriétés, justifiant l'emploi des méthodes de simulations, sont néanmoins de moindre importance pour le contrôle de la convergence d'une simulation donnée, où les propriétés de la réalisation de la chaîne ($X_n = x_n$) sont les seules caractéristiques d'intérêt. On doit ainsi considérer la différence entre l'*analyse probabiliste*, qui décrit le comportement moyen des réalisations, et l'*inférence statistique* qui traite de la réalisation observée, résultant de l'algorithme de simulation.

L'impossibilité d'appliquer directement les théorèmes de convergence classiques (loi des grands nombres, théorème de la limite centrale) à l'échantillon (ou trajectoire) $(X_0 = x_0, \dots, X_n = x_n)$ est due à la fois à la structure de dépendance markovienne entre les observations et à la non-stationnarité de la suite (à moins de simuler directement X_0 selon π , ce qui est particulièrement rare en pratique). Par conséquent, on considère ici que la chaîne démarre en un point $X_0 = x_0 \in \mathcal{X}$ dont la distribution n'est pas la distribution stationnaire.

Etant donnée une réalisation de la chaîne $(X_0 = x_0, \dots, X_n = x_n)$, on examine le comportement asymptotique des sommes partielles :

$$S_n(h) = \frac{1}{n} \sum_{i=0}^n h(X_i), \quad (3.17)$$

où n tend vers l'infini.

Théorème 3.3.5 *Si la chaîne de Markov (X_n) est ergodique et récurrente de Harris, pour tout $h \in L^1(\pi) = \{h, \int_{\mathcal{X}} |h(x)| \pi(dx) < \infty\}$, on a pour toute distribution initiale μ :*

$$S_n(h) \rightarrow \mathbb{E}_{\pi}(h), \quad \mathbb{P}_{\mu}\text{-presque sûrement.} \quad (3.18)$$

Enfin, on peut démontrer le théorème de la limite centrale suivant, dont les conditions sont aisément remplies par la plupart des algorithmes MCMC.

Théorème 3.3.6 *On suppose que (X_n) est ergodique et récurrente de Harris, de distribution stationnaire π . Soit h une fonction d -dimensionnelle sur \mathcal{X} , de la forme $h = g - Kg$, pour une fonction $g \in L^2(\pi) = \{g, \int_{\mathcal{X}} \|g(x)\|^2 \pi(dx) < \infty\}$. Pour toute distribution initiale μ , $S_n(h)$ vérifie :*

$$\frac{1}{\sqrt{n}} \left(\sum_{i=1}^n h(X_i) - \mathbb{E}_{\pi}[h] \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma),$$

où $\xrightarrow{\mathcal{L}}$ désigne la convergence en loi et Γ est une matrice $d \times d$ définie par (avec g_i désignant la $i^{\text{ème}}$ composante de g) :

$$\Gamma_{i,j} = \pi(K(g_i g_j) - (Kg_i)(Kg_j)) = \pi(h_i g_j + h_j g_i - h_i h_j)$$

Enfin, il existe d'autres conditions de convergence, comme les *conditions de dérive* et la *condition d' α -mixing* (mélange), qui permettent, entre autre, d'établir des théorèmes de la limite centrale. Elles sont détaillées dans les notes du chapitre 6 de [Robert and Casella, 2004].

Nous terminons cette section avec un mot sur la covariance dans les chaînes de Markov.

3.3.5 La covariance dans les chaînes de Markov

Une application de l'inégalité de Chebychev montre que la convergence d'une moyenne de variables aléatoires issues d'une chaîne de Markov peut être liée au comportement des covariances le long de la chaîne, une condition suffisante pour la convergence en probabilité étant que ces covariances tendent vers zero.

On a en particulier le résultat suivant :

Théorème 3.3.7 *Si la chaîne (X_n) , à valeurs réelles, est positive et apériodique, avec $\text{var}(X_n) < \infty$, alors $\text{cov}(X_0, X_n)$ converge vers 0.*

Munis de ces notions théoriques, nous pouvons maintenant aborder l'algorithme de Metropolis-Hastings et en établir les principaux résultats de convergence.

3.4 L'algorithme de Metropolis-Hastings

L'idée d'employer une chaîne de Markov de distribution limite π pour générer un échantillon de cette loi peut sembler complexe en regard des méthodes de Monte-Carlo standard présentées en section 3.2. En effet, les propriétés de convergence asymptotique reposent sur une théorie plus complexe qu'une simple loi des grands nombres, puisque nous générons des dépendances dans l'échantillon qui vont ralentir la convergence de l'approximation de (3.3) (page 86). Ainsi, le nombre d'itérations nécessaires à obtenir une approximation satisfaisante semble a priori bien plus important que dans le cadre des méthodes de Monte-Carlo standard. Néanmoins, l'emploi d'une chaîne (X_n) produite par un algorithme MCMC de distribution stationnaire π est fondamentalement identique à l'emploi d'un échantillon i.i.d. de π dans le sens où le théorème 3.3.5

garantit la convergence presque sûre de la moyenne empirique :

$$S_n(h) = \frac{1}{n} \sum_{i=1}^n h(X_i) \quad (3.19)$$

vers la quantité $\mathbb{E}_\pi[h(X)]$.

Malgré leur complexité, l'emploi de telles méthodes est justifié lorsque la simulation directe selon la loi cible est impossible à réaliser ou présente un coût informatique trop élevé, en raison de la dimension et/ou de sa structure, comme c'est le cas pour la simulation selon (3.1) (page 85), lorsque l'opérateur F est non-linéaire.

Nous introduisons maintenant l'algorithme de Metropolis-Hastings(MH). Dans toute cette section, sauf mention contraire, $x \in \mathcal{X}$ désignera l'état courant de la chaîne et $y \in \mathcal{X}$ la transition proposée.

3.4.1 Définition

L'algorithme MH nécessite le choix d'une densité conditionnelle dite *instrumentale* ou *de transition* q . Le rôle de cette densité $q = q(\cdot|x)$ consiste à proposer les éventuelles transitions de la chaîne de Markov (X_n) . L'algorithme MH peut être implémenté en pratique lorsque $q(\cdot|x)$ est aisée à simuler et est connue explicitement (à une constante indépendante de x près) ou bien *symétrique*, c'est-à-dire telle que $q(y|x) = q(x|y)$. Concernant la densité cible π , le rapport :

$$\frac{\pi(y)}{q(y|x)}$$

doit être connu à une constante indépendante de x près.

Une transition de l'état x à l'état y est acceptée avec la probabilité $\alpha(x, y)$ définie par :

$$\alpha(x, y) = \begin{cases} \min\left(1, \frac{\pi(y)}{\pi(x)} \frac{q(x|y)}{q(y|x)}\right) & \text{si } \pi(x)q(y|x) > 0, \\ 1 & \text{sinon.} \end{cases} \quad (3.20)$$

dans le cas contraire, la transition est rejetée et la chaîne de Markov reste en x . Le noyau de transition K de cette chaîne, pour $(x, A) \in \mathcal{X} \times \mathcal{B}(\mathcal{X})$ prend alors la forme suivante :

$$K(x, A) = \int_A \alpha(x, y)q(y|x)dy + \mathbb{1}_A(x) \int_{\mathcal{X}} (1 - \alpha(x, y))q(y|x)dy. \quad (3.21)$$

Cet algorithme accepte toujours le nouvel état y lorsque la quantité $\frac{\pi(y)}{q(y|x)}$ est plus grande que $\frac{\pi(x)}{q(x|y)}$. Aussi, lorsque q est symétrique, la probabilité d'acceptation ne dépend que du rapport $\frac{\pi(y)}{\pi(x)}$.

Bien que l'algorithme ainsi défini soit générique, défini pour tout π et q , il est nécessaire d'imposer des conditions de régularités minimales sur π et q pour que π soit la distribution limite de la chaîne ainsi produite. Par exemple, si \mathcal{E} , le support de π , est *déconnecté* (par exemple, si $\mathcal{X} = \mathbb{R}$ et \mathcal{E} est une réunion d'intervalles disjoints) : l'algorithme MH peut ne pas converger. Il faut alors montrer la convergence sur chaque

composante connectée de \mathcal{E} et que le noyau (3.21) permet de les relier. Pour montrer que π est la loi limite de (X_n) , définie par son noyau (3.21), on peut montrer qu'il satisfait la condition de balance détaillée (3.16) :

Théorème 3.4.1 *Soit (X_n) une chaîne de Markov de noyau (3.21). Pour toute distribution conditionnelle q dont le support contient \mathcal{E} ,*

- i) le noyau de la chaîne satisfait la condition de balance détaillée avec π ;*
- ii) π est la distribution stationnaire de la chaîne.*

La stationnarité de π est ainsi établie pour la plupart des distributions conditionnelles q , ce qui induit l'universalité des algorithmes MH. Sauf mention contraire, (X_n) désigne dorénavant une chaîne de noyau (3.21).

3.4.2 Propriétés de convergence

Montrer que (X_n) converge effectivement vers π et que (3.19) converge vers \mathcal{I} , nécessite d'appliquer la théorie abordée dans la section précédente.

Puisque (X_n) admet par construction une distribution invariante π , si elle est aussi Harris récurrente et apériodique (définition 3.3.11 et partie 3.3.2), alors le théorème 3.3.5 s'applique et assure la convergence de (3.19) vers \mathcal{I} . Une condition suffisante pour que (X_n) soit apériodique est que la probabilité d'événements tels que $\{X_{n+1} = X_n\}$ soit non nulle, c'est-à-dire :

$$\mathbb{P}(\pi(x)q(y|x) \leq \pi(y)q(x|y)) < 1 \quad (3.22)$$

La propriété d'*irréductibilité* de (X_n) provient de conditions suffisantes telles que la positivité de q :

$$q(y|x) > 0 \text{ pour tout } (x, y) \in \mathcal{E} \times \mathcal{E}, \quad (3.23)$$

puisqu'elle implique que chaque ensemble de \mathcal{E} de mesure de Lebesgue positive peut être rejoint en une seule itération. Puisque la distribution π est la mesure invariante de la chaîne, la chaîne est positive (définition 3.3.12) et la proposition 3.3.2 implique que la chaîne est récurrente. On établit alors le résultat suivant :

Lemme 3.4.1 *Si la chaîne (X_n) est π -irréductible, elle est Harris récurrente.*

On en déduit le résultat de convergence suivant :

Théorème 3.4.2 *Si la distribution instrumentale q satisfait (3.22) et (3.23), on a*

- i) Si $h \in L^1(\pi)$, alors*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n h(X_i) = \int h(x)\pi(x)dx, \text{ } \pi\text{-presque sûrement.}$$

- ii) De plus, si (X_n) est apériodique, alors*

$$\lim_{n \rightarrow \infty} \left\| \int K^n(x, \cdot) \mu(dx) - \pi \right\|_{TV} = 0$$

pour toute distribution μ .

On mentionne également le lemme suivant proposant une variante des conditions (3.22 et (3.23) impliquant l'irréductibilité et l'apériodicité de la chaîne :

Lemme 3.4.2 *S'il existe des nombres positifs ε et δ tels que :*

$$q(y|x) > \varepsilon \text{ si } |x - y| < \delta, \quad (3.24)$$

alors la chaîne (X_n) est irréductible et apériodique.

Par suite, si q vérifie (3.24), alors les conclusions du théorème 3.4.2 s'appliquent. L'un des aspects les plus fascinants de l'algorithme MH est son universalité, c'est-à-dire le fait qu'une distribution instrumentale q arbitraire de support \mathcal{E} permet la simulation d'une distribution π de même support. Cependant, cette universalité est effective uniquement formellement si la distribution instrumentale q simule rarement des points dans les régions les plus importantes de \mathcal{E} , au sens où elles portent l'essentiel de la masse de π .

Nous présentons maintenant quelques exemples pour le choix de q , basés sur les implémentations les plus courantes de l'algorithme MH.

3.4.3 Echantillonneur indépendant

Cette méthode apparaît comme une généralisation directe de la méthode de simulation d'acceptation-rejet (cf. [Robert and Casella, 2004], chapitre 2). Sa distribution instrumentale est indépendante de x et s'écrit donc $q(y)$. Sa probabilité de transition s'écrit :

$$\alpha(x, y) = \begin{cases} \min \left(1, \frac{\pi(y)q(x)}{\pi(x)q(y)} \right) & \text{si } \pi(x)q(y) > 0, \\ 1 & \text{sinon.} \end{cases} \quad (3.25)$$

Le noyau de transition K de cette chaîne, pour $(x, A) \in \mathcal{X} \times \mathcal{B}(\mathcal{X})$ prend alors la forme suivante :

$$K(x, A) = \int_A \alpha(x, y)q(y)dy + \mathbf{1}_A(x) \int_{\mathcal{X}} (1 - \alpha(x, y))q(y)dy. \quad (3.26)$$

Bien que les y soient générés indépendamment de x , l'échantillon simulé n'est pas i.i.d. : la probabilité d'acceptation de y dépend de x .

Les propriétés de convergence de (X_n) ainsi générée vient des propriétés de la densité instrumentale q . (X_n) est irréductible et apériodique si et seulement si q est presque partout positive sur le support de π . Les propriétés de convergence géométrique et d'ergodicité uniforme sont décrites dans le résultat suivant :

Théorème 3.4.3 *L'échantillonneur indépendant produit une chaîne uniformément ergodique si il existe une constante M telle que*

$$\pi(x) \leq Mq(x), \quad \forall x \in \mathcal{X}. \quad (3.27)$$

Dans ce cas,

$$\|K^n(x, \cdot) - \pi\|_{TV} \leq 2 \left(1 - \frac{1}{M} \right)^n. \quad (3.28)$$

Le lemme suivant indique que la probabilité d'acceptation est plus grande avec cet algorithme qu'avec l'algorithme d'acceptation-rejet.

Lemme 3.4.3 *Si (3.27) est vérifiée, l'espérance du taux d'acceptation associé à l'échantillonneur indépendant est au moins $1/M$ quand la chaîne est stationnaire.*

Enfin, il est important de noter que l'échantillonneur indépendant (IMH) n'est pas universel. Il existe des cas où une proposition indépendante ne fonctionne pas de manière satisfaisante, en raison de la complexité de la loi cible ; les propositions sont rejetées beaucoup trop souvent. Puisque l'objectif principal des algorithmes MCMC est de fournir une méthode de simulation simple, il est difficile d'imaginer passer beaucoup de temps à construire la loi instrumentale. Ceci est particulièrement le cas en grande dimension, lorsque les caractéristiques de la loi cible sont la plupart du temps impossibles à identifier a priori. Dans le cadre des problèmes inverses, la construction d'une bonne loi instrumentale pour des propositions indépendantes suppose finalement que le problème soit déjà résolu. Ce problème est une limitation importante de l'échantillonneur indépendant. Il peut être vu comme une proposition *globale* alors que l'on a parfois besoin de propositions plus *locales* qui sont moins sensibles à la loi cible, comme celles présentées dans la section 3.4.4.

Une alternative est de simuler indépendamment une ou plusieurs composantes à la fois, on parlera alors d'échantillonneur séquentiellement indépendant. D'autres possibilités développées plus loin passent par l'emploi de l'échantillonneur de Langevin (3.33) ou bien de méthodes adaptatives (3.34). Cette dernière solution est cependant délicate à la fois du point de vue théorique (*A-t-on ergodicité ?*) et algorithmique (*Comment optimiser la procédure d'adaptation ?*).

3.4.4 Marches aléatoires

Une approche naturelle pour la construction pratique d'algorithmes MH est de perturber l'état courant de la chaîne pour générer l'état suivant, autrement dit, de considérer une exploration *locale* du voisinage de l'état courant de la chaîne. Un premier choix est de simuler y selon :

$$y = x + \varepsilon,$$

où ε est une perturbation aléatoire de distribution q , indépendante de x . Ici, $q(y|x) = q(y - x)$. La chaîne de Markov associée à q est une marche aléatoire (voir définition 3.3.3) sur \mathcal{E} . (X_n) désigne dans la suite une marche aléatoire.

Les résultats de convergence s'appliquent naturellement dans ce cas particulier. Suivant le lemme 3.4.2, si $q(y - x)$ est positive dans un voisinage de zero, (X_n) est irréductible et apériodique, et donc ergodique.

Les distributions q généralement employées dans ce cas sont la distribution uniforme sur des sphères centrées à l'origine, la distribution gaussienne, et la distribution de Student.

Le choix d'une fonction q symétrique (telle que $q(x) = q(-x)$) conduit à l'expression suivante de la probabilité d'acceptation :

$$\alpha(x, y) = \begin{cases} \min \left(1, \frac{\pi(y)}{\pi(x)} \right) & \text{si } \pi(x)q(y) > 0, \\ 1 & \text{sinon.} \end{cases} \quad (3.29)$$

Elle ne dépend ainsi que du rapport $\pi(y)/\pi(x)$.

Le noyau de transition K de cette chaîne, pour $(x, A) \in \mathcal{X} \times \mathcal{B}(\mathcal{X})$ prend alors la forme suivante :

$$K(x, A) = \int_A \alpha(x, y)q(y - x)dy + \mathbb{1}_A(x) \int_{\mathcal{X}} (1 - \alpha(x, y))q(y - x)dy. \quad (3.30)$$

En dépit de sa simplicité et de ses propriétés naturelles, l'algorithme MH de marche aléatoire (RWMH) ne bénéficie pas de propriétés de convergence uniforme. Ceci n'est pas forcément une surprise compte tenu du caractère *local* de la proposition y , centrée sur l'état courant de la chaîne.

Il est cependant possible de définir des conditions suffisantes pour la convergence géométrique.

Théorème 3.4.4 *Supposons que π est symétrique et log-concave, c'est-à-dire s'il existe $\beta > 0$ et x_1 tels que :*

$$\log(\pi(x)) - \log(\pi(y)) > \beta|y - x|,$$

pour $y < x < -x_1$ et $x_1 < x < y$.

Si q est positive et symétrique, la chaîne (X_n) est géométriquement ergodique. Si π n'est pas symétrique, une condition suffisante pour l'ergodicité géométrique est que $q(t)$ soit bornée par $b \exp(-\alpha|t|)$ pour une constante b suffisamment grande.

Nous renvoyons à [Robert and Casella, 2004], chapitre 7 et aux références incluses pour plus de détails.

3.4.5 Echantillonneurs autorégressifs

Tierney ([Tierney, 1994]) a proposé une modification du précédent algorithme avec une distribution instrumentale de la forme $q(y - a - b(x - a))$, c'est-à-dire :

$$y = a + b(x - a) + \varepsilon, \quad \varepsilon \sim q$$

Cette représentation autorégressive peut être vue comme un intermédiaire entre l'échantillonneur indépendant ($b = 0$) et la marche aléatoire ($b = 1$). De plus, lorsque $b < 0$, X_n et X_{n+1} sont corrélés négativement et cela peut permettre une exploration rapide de \mathcal{E} si le centre de symétrie a est bien choisi.

3.4.6 Echantillonneur de Langevin

Des alternatives à l'algorithme RWMH proviennent de la *théorie de la diffusion*. L'idée de base de cette approche est de chercher une *équation de diffusion* (ou *équation différentielle stochastique*) qui produit une *diffusion* de distribution stationnaire π , puis de discrétiser le processus pour implémenter la méthode. Notamment, la *diffusion de Langevin* est définie par l'équation différentielle stochastique suivante :

$$dL(t) = \frac{1}{2} \log \pi(L(t))dt + dW_t, \quad (3.31)$$

où W_t est le mouvement brownien standard. L'implémentation de l'algorithme de diffusion comprend une étape de discrétisation où (3.31) est remplacé par la transition :

$$y = x + \frac{h^2}{2} \nabla \log(\pi(x)) + h\varepsilon, \quad (3.32)$$

où $\varepsilon \sim \mathcal{N}(0, h^2 I_d)$ et h est le pas de discrétisation. Bien que cette discrétisation vise à reproduire la convergence de la marche aléatoire vers le mouvement brownien, le comportement de la chaîne de Markov définie par (3.32) peut être très différent de celui du processus de diffusion (3.31). Dans certain cas, elle peut même être transiente.

Un moyen de corriger ce comportement négatif est de voir (3.32) comme une distribution instrumentale dans le cadre de l'algorithme MH, c'est-à-dire

$$q(x, y) \sim \mathcal{N}\left(x + \frac{h^2}{2} \nabla \log(\pi(x)), h^2 I_d\right), \quad (3.33)$$

L'algorithme MH correspondant (LMH) n'est pas forcément supérieur au RWMH. On peut montrer que la chaîne n'est pas géométriquement ergodique quand $\nabla \log \pi(x)$ tend vers zero en l'infini. En revanche, l'ergodicité de cette chaîne est assurée.

On peut noter l'analogie entre l'équation (3.32) et un développement de Taylor du premier ordre. A ce titre, on peut trouver dans [Stramer and Tweedie, 1999] une version faisant intervenir le terme d'ordre 2 du développement.

Enfin, le choix du pas de discrétisation h dans (3.32) nécessite une attention particulière. Un mauvais choix peut induire des comportements très erratiques. Pour plus de détails sur le choix de ce paramètre, nous renvoyons vers [Roberts and Rosenthal, 1998] et [Stramer and Tweedie, 1999].

3.4.7 Schémas adaptatifs

Compte tenu de la largeur du champ d'application des méthodes MCMC, l'existence d'un échantillonneur *générique*, que l'on pourrait employer systématiquement, est irréaliste. En grande dimension et quand le support de la loi cible est déconnecté, les échantillonneurs les plus généraux comme les marches aléatoires ne sont pas satisfaisants puisqu'ils sont trop lents à parcourir l'espace d'intérêt.

Pour autant, les performances d'algorithmes comme la marche aléatoire apportent de l'information à propos de la distribution d'intérêt et peuvent donc être incorporés dans la construction d'algorithmes plus performants. Alors qu'il semble naturel de penser que l'information fournie par les premières itérations de l'algorithme devrait être exploitée pour les itérations suivantes, ce constat implique qu'il ne s'agit plus une chaîne de Markov. Par conséquent, les résultats de convergence classiques ne s'appliquent plus et la validité de tels algorithmes est mise en cause. De plus, de tels algorithmes peuvent en pratique dégénérer en raison d'une décroissance trop rapide de la variation de leur loi instrumentale.

Bien que la chaîne de Markov puisse converger *en loi* vers la loi cible (lorsque l'on emploie un schéma d'adaptation homogène en temps), un très grand nombre de simulations n'est généralement pas suffisant pour atteindre une approximation satisfaisante de la loi cible.

Le message principal est donc qu'il ne faut pas adapter constamment la loi instrumentale par rapport aux performances passées de la chaîne. Ou bien l'adaptation cesse au-delà d'une période de "*burn-in*", ou bien le schéma adaptatif doit être évalué théoriquement, et ses propriétés de convergence démontrées.

Le premier schéma de ce type est apparu dans [Holden, 1998]. Une application de sa méthode au problème de calage d'historique en ingénierie de réservoir est présentée dans [Holden et al., 2002]. On peut citer également [Haario et al., 2001], qui propose d'optimiser les propositions gaussiennes dans une marche aléatoire en estimant leurs paramètres le long de la chaîne. Les mêmes auteurs en proposent une version adaptée aux grandes dimensions dans [Haario et al., 2003], et l'appliquent au problème inverse de la télédétection des composants de l'atmosphère dans [Haario et al., 2004]. Dans [Roberts and Rosenthal, 2006], les auteurs présentent un certain nombre d'exemples de schémas adaptatifs.

Nous partons dans un premier temps de l'algorithme de [Haario et al., 2001] qui nous permettra d'introduire le cadre général proposé par Andrieu et Robert ([C. Andrieu and C. Robert, 2001]), basé sur l'*optimisation stochastique* et l'algorithme de *Robbins-Monro*, voir [Robbins and Monro, 1951]. Dans ce cadre, les principaux résultats de convergence sont établis dans [Andrieu and Moulines, 2003].

Cadre général

L'algorithme de [Haario et al., 2001] est basé sur une marche aléatoire à accroissements symétriques où un choix classique pour la distribution q est une loi du type $\mathcal{N}(0, \Gamma)$. En particulier, le choix de Γ est généralement délicat. Les auteurs ont alors proposé d'estimer Γ le long de la chaîne. Dans cet algorithme, X_{n+1} est proposé selon $q_{\theta_n}(\cdot | X_n) = \mathcal{N}(X_n, \Gamma_n)$, où $\theta = (\mu, \Gamma) \in \Theta = \mathbb{R}^d \times \mathcal{C}_+^d$, où \mathcal{C}_+^d est le cône des matrices positives $d \times d$, et d la dimension de \mathcal{X} . On considère également une suite $\{\gamma_n\}$ non-croissante de pas positifs, telle que $\sum_{n=1}^{\infty} \gamma_n = \infty$ et $\sum_{n=1}^{\infty} \gamma_n^{1+\delta} < \infty$ pour un certain $\delta > 0$. En pratique, on emploiera la suite : $\gamma_n = 1/n$, qui revient à calculer les moyennes et covariances empiriques. L'algorithme d'estimation des paramètres se résume de la manière suivante :

$$\begin{aligned} \mu_{n+1} &= \mu_n + \gamma_{n+1} (X_{n+1} - \mu_n), n \geq 0, \\ \Gamma_{n+1} &= \Gamma_n + \gamma_{n+1} \left((X_{n+1} - \mu_n)(X_{n+1} - \mu_n)^T - \Gamma_n \right), \end{aligned} \quad (3.34)$$

Il a été réalisé dans [C. Andrieu and C. Robert, 2001] qu'un tel schéma est un cas particulier d'un cadre plus général.

Plus précisément, on définit $H : \Theta \times \mathcal{X} \rightarrow \mathbb{R}^d \times \mathbb{R}^{d \times d}$ telle que

$$H(\theta, x) = (x - \mu, (x - \mu)(x - \mu)^T - \Gamma). \quad (3.35)$$

Avec cette notation, la récursion (3.34) s'écrit

$$\theta_{n+1} = \theta_n + \gamma_{n+1} H(\theta_n, X_{n+1}), \quad n \geq 0, \quad (3.36)$$

avec $X_{n+1} \sim \mathcal{N}(X_n, \Gamma_n)$. Ce type de récursion est au coeur de la plupart des algorithmes d'approximation stochastique classiques, sous le nom d'algorithme de Robbins-Monro

(voir [Robert and Casella, 2004] pour une description de l'algorithme). Cet algorithme est conçu pour résoudre des équations du type $h(\theta) = 0$, où $\theta \rightarrow h(\theta)$ est le *champ moyen* (*mean field*), défini par

$$h(\theta) = \int_{\mathcal{X}} H(\theta, x) \pi(dx). \quad (3.37)$$

Dans le cas présent, si l'on suppose que $\int_{\mathcal{X}} |x|^2 \pi(dx) < \infty$, on peut facilement vérifier que

$$h(\theta) = \int_{\mathcal{X}} H(\theta, x) \pi(dx) = (\mu_{\pi} - \mu, (\mu_{\pi} - \mu)(\mu_{\pi} - \mu)^t + \Gamma_{\pi} - \Gamma),$$

où

$$\mu_{\pi} = \int_{\mathcal{X}} x \pi(dx) \text{ et } \Gamma_{\pi} = \int_{\mathcal{X}} (x - \mu_{\pi})(x - \mu_{\pi})^t \pi(dx)$$

On peut réécrire (3.36) de la manière suivante

$$\theta_{n+1} = \theta_n + \gamma_{n+1} h(\theta_n) + \gamma_{n+1} \xi_{n+1}, \quad (3.38)$$

où $\{\xi_{n+1} = H(\theta_{n-1}, X_n) - h(\theta_{n-1}), n \geq 1\}$ est généralement appelé *le bruit*. La théorie générale de l'approximation stochastique fournit les conditions sous lesquelles cette récursion converge vers l'ensemble $\{\theta \in \Theta, h(\theta) = 0\}$. La chaîne de Markov inhomogène définie par un tel algorithme sera notée (X_n) .

Résultats de convergence

Les auteurs de [Andrieu and Moulines, 2003] imposent un certain nombre de conditions afin d'établir la convergence des algorithmes adaptatifs.

En premier lieu, ils introduisent une famille paramétrée de noyaux de transition sur \mathcal{X} , $\{K_{\theta}, \theta \in \Theta\}$, indexée par un paramètre de dimension finie $\theta \in \Theta \subset \mathbb{R}^{d_{\theta}}$, un ensemble ouvert. Pour tout $\theta \in \Theta$, on suppose que K_{θ} est π -irréductible et que $\pi K_{\theta} = \pi$, c'est-à-dire π est la loi invariante de K_{θ} .

Pour $W : \mathcal{X} \rightarrow [1, \infty)$ et $f(\theta, x) : \Theta \times \mathcal{X} \rightarrow \mathbb{R}$, que l'on notera $f_{\theta}(x)$, une fonction mesurable, on définit

$$\|f_{\theta}\|_W = \sup_{x \in \mathcal{X}} \frac{|f_{\theta}(x)|}{W(x)} \text{ et } \mathcal{L}_W = \{f_{\theta} : \|f_{\theta}\|_W < \infty\}. \quad (3.39)$$

On dira que la famille de fonctions $\{f_{\theta} : \Theta \times \mathcal{X} \rightarrow \mathbb{R}, \theta \in \Theta\}$ est *W-Lipschitz* si pour tout sous-ensemble compact \mathcal{K} de Θ ,

$$\sup_{\theta \in \mathcal{K}} \|f_{\theta}\|_W < \infty \text{ et } \sup_{(\theta, \theta') \in \mathcal{K} \times \mathcal{K}, \theta \neq \theta'} |\theta - \theta'|^{-1} \|f_{\theta} - f_{\theta'}\|_W < \infty.$$

On définit alors les trois conditions suivantes :

(A-1) Pour tout $\theta \in \Theta$, il existe une fonction $V : \mathcal{X} \rightarrow [1, \infty)$ telle que $\sup_{x \in \mathcal{X}} V(x) < \infty$ et telle que, pour tout compact $\mathcal{K} \subset \Theta$,

- i) *Condition minorante*. Il existe $C \in \mathcal{B}(\mathcal{X})$, $\varepsilon > 0$ et une mesure de probabilité ν (tous trois dépendants de \mathcal{K}) tels que $\nu(C) > 0$ et pour tout $A \in \mathcal{B}(\mathcal{X})$ et $\theta, x \in \mathcal{K} \times C$

$$K_{\theta}(x, A) > \varepsilon \nu(A)$$

- ii) *Condition de dérive.* Il existe des constantes $\lambda \in [0, 1)$, $b \in (0, \infty)$, qui vérifient

$$K_\theta V(x) \leq \begin{cases} \lambda V(x) & x \notin C, \\ b & x \in C \end{cases}$$

pour tout $\theta \in \mathcal{K}$.

- (A-2) Pour tout compact $\mathcal{K} \subset \Theta$ et tout $r \in [0, 1]$, il existe une constante c telle que, pour tout $(\theta, \theta') \in \mathcal{K} \times \mathcal{K}$ et $f \in \mathcal{L}_{V^r}$,

$$\|K_\theta f - K_{\theta'} f\|_{V^r} \leq c \|f\|_{V^r} |\theta - \theta'|,$$

où V est donnée en (A1).

- (A-3) $\{H_\theta, \theta \in \Theta\}$ est V^β -Lipschitz pour $\beta \in [0, 1/2]$ et V donnée en (A1).

La condition (A-1) a été largement étudiée ces dix dernières années pour prouver l'ergodicité géométrique des algorithmes MH. Typiquement, la fonction V est une fonction potentiel qui dépend de la loi cible π . Ces assertions permettent de dériver une loi des grands nombres :

Théorème 3.4.5 *Soit $\{\gamma_n\}$ une suite positive non-croissante telle que $\sum_{k=1}^{\infty} k^{-1} \gamma_k < \infty$. On suppose (A1-3) et soit $f : \mathcal{X} \rightarrow \mathbb{R}$ une fonction telle que $\|f\|_{V^\alpha} < \infty$ pour $\alpha \in [0, 1 - \beta)$, avec β et V définis respectivement en (A-3) et (A-1). On suppose de plus que le nombre de mise à jour des paramètres est presque sûrement fini. Alors,*

$$\frac{1}{n} \sum_{i=1}^n \left(f(X_i) - \int_{\mathcal{X}} f(x) \pi(dx) \right) \xrightarrow{p.s.} 0. \quad (3.40)$$

En supposant de plus la convergence de θ_n , on peut dériver un théorème de la limite centrale.

Théorème 3.4.6 *On se place sous les hypothèses de 3.4.5. On note, pour $\theta \in \Theta$,*

$$\sigma^2(\theta, f) = \pi \left[(\hat{f}_\theta - K_\theta \hat{f}_\theta)^2 \right] \text{ avec } \hat{f}_\theta = \sum_{i=0}^{\infty} [K_\theta^i f - \pi(f)].$$

On suppose de plus qu'il existe une variable aléatoire $\theta_\infty \in \Theta$, telle que $\int_{\mathcal{X}} \widehat{f_{\theta_\infty}}^2 \pi(dx) < \infty$ et $\int_{\mathcal{X}} K_{\theta_\infty} \widehat{f_{\theta_\infty}}^2 \pi(dx) < \infty$ presque sûrement, et

$$\limsup_{n \rightarrow \infty} |\theta_n - \theta_\infty| = 0, \text{ p.s.}$$

Si de plus, $\sigma^2(\theta_\infty, f) > 0$, p.s., alors,

$$\frac{1}{\sqrt{n} \sigma(\theta_\infty, f)} \sum_{i=1}^n \left(f(X_i) - \int_{\mathcal{X}} f(x) \pi(dx) \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1). \quad (3.41)$$

Des conditions supplémentaires sur π et l'hypothèse que $\theta_\infty = \theta_\pi$ permettent d'appliquer le théorème 3.4.6 à l'algorithme (3.34).

Cependant, cette méthode ne paraît pas forcément adaptée à notre problème dans la mesure où elle va cibler la loi gaussienne multivariée la plus proche de la loi a posteriori. En particulier, elle semble mal adaptée à l'inférence de lois multi-modales, comme nous le montrons en section 3.5.

3.4.8 Recuit et trempe simulés

L'algorithme de recuit simulé a été introduit par [Metropolis et al., 1953], puis généralisé par [Kirkpatrick et al., 1983] pour des problèmes d'optimisation. Il peut s'appliquer à la fois à des problèmes d'optimisation globale et de simulation (cf. [Robert and Casella, 2004] et sa bibliographie). Sa dénomination vient d'une analogie avec la métallurgie. Un métal est une forme cristalline au sein de laquelle des atomes peuvent venir se loger ; un métal obtenu par une lente décroissance de la température après sa mise en solution (chauffage) sera plus solide qu'un métal obtenu en la réduisant rapidement. Au contraire, la trempe est un refroidissement brutal de la pièce de métal qui a pour objectif de figer la structure cristalline obtenue lors de la mise en solution. L'algorithme de simulation analogue (*simulated tempering*) été introduit indépendamment dans [Marinari and Parisi, 1992] et [Geyer and Thompson, 1995]. Ces méthodes, comme d'ailleurs les méthodes McMC, ont été initialement développées par des physiciens pour la simulation de mécanique statistique présentant de nombreux états d'énergie minimale. Au contraire des méthodes McMC, leur terminologie reste fortement liée aux modèles de mécanique statistique dont la simulation est l'objet de ces méthodes. Elles visent en particulier à générer des échantillons issus d'une distribution de Gibbs :

Définition 3.4.1 *Un champ aléatoire X , à valeur dans \mathcal{X} , est un champ de Gibbs d'énergie E , s'il est régi par la loi de probabilité de densité (par rapport à la mesure de Lebesgue) :*

$$f(x) = \frac{1}{\mathcal{Z}} e^{-E(x)}, \quad (3.42)$$

$$\mathcal{Z} = \int_{\mathcal{X}} e^{-E(x)} dx, \quad (3.43)$$

appelée *densité de Gibbs*.

La constante \mathcal{Z} est généralement incalculable en raison de la dimension de X , c'est pourquoi les méthodes McMC et leurs variantes que sont le recuit et la trempe simulés sont particulièrement adaptés à leur simulation. On notera que la distribution a posteriori du problème inverse (3.1) est une distribution de Gibbs (p. 85).

L'idée fondamentale de ces algorithmes est qu'un changement d'échelle, appelée *température*, permet des mouvements plus rapides sur la surface de la fonction E à minimiser/explorer que les méthodes d'optimisations déterministes comme la méthode du gradient, ou que les méthodes McMC classiques. Ainsi, ce changement d'échelle permet d'éviter, en théorie, de rester bloqué dans le bassin d'attraction d'un minimum/mode local.

Recuit simulé

Nous décrivons dans un premier temps l'algorithme de recuit simulé. Etant donné une suite de températures croissantes positives (T_n) , une chaîne $X_1 = x_1, X_2 = x_2, \dots$ est générée à partir des densités de Gibbs suivantes :

$$\pi_i(x) \propto \exp(-E(x)/T_i).$$

Alors que T_i décroît, les valeurs simulées selon cette densité se concentrent dans un voisinage toujours plus restreint du minimum de E . En effet, T_i diminuant, la densité cible va être de moins en moins "plate" : à haute température, elle aura une forme très dilatée, ce qui permettra à la chaîne générée de parcourir aisément le paysage énergétique, tandis que lorsque T_i tendra vers 0, la densité cible se rapprochera d'une masse de Dirac au minimum de E .

Cet algorithme peut être vu comme une version non homogène de l'algorithme MH. En effet, dans sa version la plus courante, (T_n) décroît à chaque itération, le noyau de la chaîne varie donc au cours du temps. Les résultats de convergence de la section 3.3 ne s'appliquent donc pas dans ce cadre. L'algorithme se présente de la manière suivante :

1. générer y selon une distribution instrumentale de densité $q(|y - x_i|)$,
2. $x_{i+1} = y$ avec probabilité $\alpha = \inf(\exp((E(y) - E(x))/T_i), 1)$, $x_{i+1} = x_i$ sinon,
3. T prend la valeur T_{i+1} .

On applique généralement des règles heuristiques pour assurer la validité de cet algorithme : il est nécessaire que la température de départ soit suffisamment élevée et sa décroissance suffisamment lente. Un résultat de convergence existe, sous conditions sur E , avec l'hypothèse que la température décroisse en $1/\log(n)$. Dans la pratique néanmoins, on emploie généralement une suite à décroissance géométrique.

Trempe simulée

Le principe de l'algorithme de trempe simulée s'inspire du recuit simulé dans le sens où l'on va également considérer des distributions de Gibbs dilatées par un paramètre de température T . En revanche le but de cet algorithme n'est pas tant de minimiser l'énergie d'un système que d'échantillonner selon une loi cible π . Le principe est d'employer une suite finie de températures et de distributions de Gibbs associées, et d'autoriser la chaîne à changer de niveau de température selon une certaine probabilité. Ceci permettra à la chaîne de repasser par des températures plus élevées afin d'échapper à l'attraction d'un mode local, résultant en une amélioration des propriétés de mélange de la chaîne.

On définit dans un premier temps une suite croissante de $(K+1)$ températures $1 = T_0 < \dots < T_K$, les densités de Gibbs associées $\pi_i(x) \propto e^{-\frac{E(x)}{T_i}}$, qui sont des versions dilatées de $\pi_0 = \pi$ pour $i > 0$, une variable auxiliaire M , à valeurs dans $\{0, \dots, K\}$, et la distribution jointe :

$$\mu(x, m) = \rho_m \pi(x), \quad \sum_{i=0}^K \rho_m = 1. \quad (3.44)$$

On définit également les probabilités p_U et p_D de proposer une transition vers la température T_{i+1} (*up*), respectivement T_{i-1} (*down*), sachant que la chaîne se trouve à la température T_i , et la probabilité de réaliser une transition classique $(1 - p_U - p_D)$.

Le principe est de simuler une chaîne (X_n, M_n) à valeur dans $\mathcal{X} \times \{0, \dots, K\}$. En notant respectivement $q_{i \rightarrow i+1}(X_{n+1}|X_n = x_n)$ et $q_{i+1 \rightarrow i}(X_{n+1}|X_n = x_n)$ les probabilités de proposition de transition vers la température supérieure et vers la température inférieure, la probabilité d'acceptation d'une transition de T_i vers T_{i+1} est alors proportionnelle

à :

$$\rho_{i \rightarrow i+1}(x_n, x_{n+1}) = \frac{p_D}{p_U} \frac{\pi_{i+1}}{\pi_i} \frac{q_{i \rightarrow i+1}(X_{n+1} = x_{n+1} | X_n = x_n)}{q_{i+1 \rightarrow i}(X_{n+1} = x_{n+1} | X_n = x_n)}, \quad (3.45)$$

Pour maintenir la condition de balance détaillée, il est alors nécessaire que $\rho_{i+1 \rightarrow i}(x_{n+1}, x_n) = 1/\rho_{i \rightarrow i+1}(x_n, x_{n+1})$ et de choisir les distributions de propositions $q_{i \rightarrow i+1}(X_{n+1} | X_n = x_n)$ et $q_{i+1 \rightarrow i}(X_{n+1} | X_n = x_n)$ en conséquence. Pour autant, il est nécessaire de noter que (3.45) dépend du rapport des constantes de normalisation de π_{i+1} et π_i . On peut passer outre cette difficulté en considérant un nombre égal de transitions de i vers $i+1$ et de $i+1$ vers i , et en considérant l'entière séquence des distributions comme une unique proposition, ce qui permet de supprimer les constantes de normalisation dans la probabilité d'acceptation, comme décrit dans [Celeux et al., 1999]. Plus précisément, à l'itération n , la séquence consiste à générer y_1 à partir de x_{n-1} , selon une transition MCMC ciblant π_1 , puis y_2 à partir de y_1 selon π_2 , jusqu'à générer y_K selon π_K ; ensuite l'on redescend l'échelle de température de la même façon, jusqu'à générer y_{2K} selon π_0 , qui sera la proposition pour l'état x_n . La probabilité d'acceptation de cette séquence est ainsi indépendante des constantes de normalisation. Ce schéma semble coûteux puisqu'il revient à se débarrasser des états de la chaîne à une température différente de T_0 . Le gain potentiel dans une meilleure exploration du support de π , autrement dit un meilleur mélange, ne nous semble pas compenser le surplus du nombre d'évaluations du problème direct ($2K$ plus important) dans la problématique du problème inverse. Néanmoins, l'exposé de cette méthode est une bonne introduction aux algorithmes de chaînes de Markov en interaction présentés en section 3.6.

3.4.9 Considérations pratiques

En pratique, à moins que X_0 ne soit simulé selon π , la chaîne requiert un certain nombre d'itérations avant d'entrer en *régime stationnaire*. Autrement dit, au-delà d'un certain *temps de chauffe* (*burn-in period*), les nouveaux états de la chaîne sont simulés approximativement selon π . On ne garde généralement dans l'échantillon que les états générés au-delà, pour l'estimation de \mathcal{I} . Par conséquent, il est nécessaire de disposer d'un critère caractérisant l'entrée de la chaîne en régime stationnaire.

Si le but de la simulation est d'étudier les propriétés de π , il n'est pas nécessaire de générer N chaînes indépendantes $(X_n^{(i)})$ ($i = 1, \dots, N$) où seulement certaines valeurs "finales" de chaque chaîne sont conservées : d'une part, le choix du nombre d'itérations pour chaque chaîne peut induire un biais, et d'autre part, cette approche implique un gaspillage considérable de simulations outre les valeurs finales sélectionnées. En d'autres termes, une unique réalisation (ou *trajectoire*) d'une chaîne de Markov est suffisante pour assurer une approximation satisfaisante de \mathcal{I} par (3.19) (p. 97) pour les fonctions h d'intérêt.

Concernant le choix de la distribution instrumentale dans l'algorithme de marche aléatoire (RWMH), on voit souvent recommandé dans la littérature d'employer des distributions dont le taux d'acceptation est proche de $1/4$ pour les modèles de grandes dimensions et égal à $1/2$ pour les modèles de dimension 1 ou 2. Cette règle heuristique est basée sur le comportement asymptotique d'un *critère d'efficacité* égal au rapport de la variance d'un estimateur basé sur un échantillon i.i.d. et la variance de l'estimateur

(3.19), c'est-à-dire

$$\frac{1}{1 + 2 \sum_{k>0} \text{corr}(X_n, X_{n+k})}$$

Le choix de la distribution de proposition est capital ; en optimisant le mécanisme de proposition, on réduit de fait le nombre de transitions nulles, c'est-à-dire, dans la problématique de l'inversion bayésienne, le nombre de simulations de l'opérateur direct. Cependant, lorsque la distribution cible a un support \mathcal{E} fortement déconnecté et/ou la dimension du problème est très grande, l'optimisation de la loi instrumentale ne suffit pas toujours pour assurer la convergence de (3.19) vers \mathcal{I} .

3.5 Application

Il existe un principal écueil dans la pratique des algorithmes MCMC pour la résolution des problèmes inverses. Les algorithmes de Metropolis-Hastings comme la marche aléatoire à accroissements symétriques (SIMH), l'échantillonneur indépendant séquentiel (c'est-à-dire composantes par composantes) où l'échantillonneur de Langevin (LMH) souffrent de mélange lent ou même de non-convergence vers le régime stationnaire, particulièrement pour des problèmes en grande dimension et/ou un nombre restreint d'itérations. Cela entraîne une mauvaise exploration de la distribution a posteriori et de longues autocorrélations le long de la chaîne, ce qui va ralentir la convergence de l'estimateur (3.19), vers \mathcal{I} . En outre, un second défaut de ces algorithmes est le faible taux d'acceptation des transitions ; il a même tendance à empirer lorsque la dimension du problème augmente. Dans le cadre des problèmes inverses en général, et du problème du calage d'historique en particulier, où l'on cherche à limiter le nombre d'évaluation de l'opérateur direct, on souhaiterait avoir un bon taux d'acceptation dans l'emploi des méthodes MCMC.

Dans cette section, nous allons illustrer ces défauts sur un problème inverse simple.

3.5.1 Le problème inverse

Dans cette section, nous allons chercher à simuler selon la loi a posteriori du problème inverse caractérisé par l'opérateur direct suivant :

$$\begin{aligned} F : \mathbb{R}^2 &\mapsto \mathbb{R} \\ X &\rightarrow 2 X_1^2 + X_2^2, \end{aligned} \tag{3.46}$$

étant donnée la loi a priori sur X :

$$X \sim \mathcal{N}\left(0, \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}\right). \tag{3.47}$$

On notera la matrice de covariance de la loi a priori C_X . On se donne une réalisation de X de référence, notée $x^* = (1.514, 1.335)$, et l'on suppose que l'on observe $d = F(x^*)$ avec une erreur $\varepsilon \sim \mathcal{N}(0, 0.5)$. L'ensemble $\{X \in \mathbb{R}^2 | F(X) = d\}$ décrit une ellipse centrée en 0. On est alors en droit de s'attendre à ce que les réalisations de la loi

a posteriori soient réparties autour de cette ellipse ; l'expression de cette loi, à une constante multiplicative près indépendante de X , est :

$$\pi(X) = P(X|d) \propto \exp \left(\|d - F(X)\|^2 + \frac{1}{2} \|X\|_{C_X^{-1}}^2 \right). \quad (3.48)$$

Bien que le problème inverse posé ici soit très simple, il est remarquable que la constante de normalisation du posterior ne puisse pas être calculée explicitement. Nous représentons la densité a posteriori selon deux points de vue différents sur la figure 3.1.

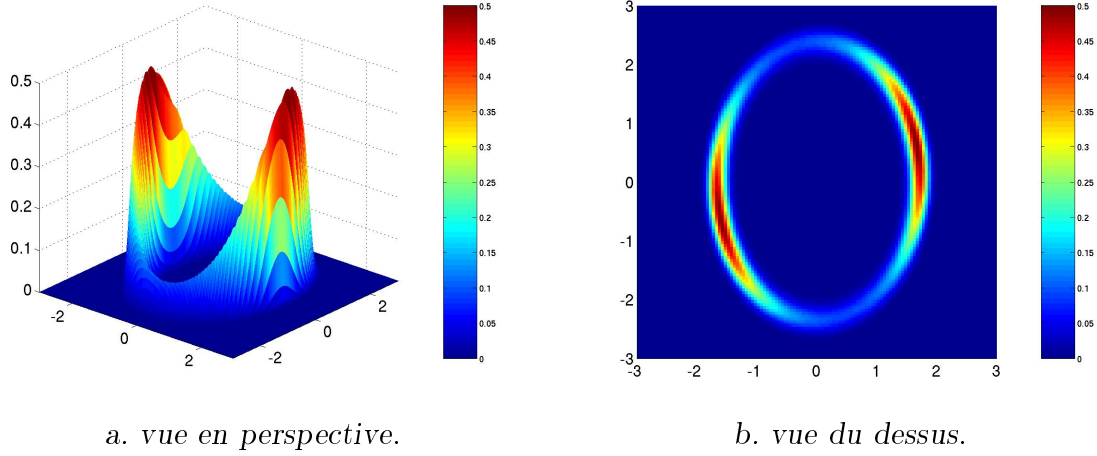


FIG. 3.1 – Densité a posteriori.

3.5.2 Comparaison des performances des algorithmes MH sur un problème inverse simple

Pour résoudre ce problème, on a implémenté 4 algorithmes MH différents :

1. l'échantillonneur indépendant (IMH) : $q(y|x) = q(y)$, où q est donné par (3.47),
2. la marche aléatoire à accroissements symétriques (SIMH) : $q(y|x) = q(|y - x|)$, où q la distribution multivariée $\mathcal{N}(0, hI)$, où I est la matrice identité et $h = 0.1$,
3. l'échantillonneur de Langevin (LMH) : $q(y|x) \sim \mathcal{N} \left(x + \frac{h^2}{2} \nabla \log(\pi(x)), h^2 I_d \right)$, où $h = 0.1$.
4. L'algorithme adaptatif de [Haario et al., 2001] (ASIMH) : $q_{\theta_k}(y|x) = \mathcal{N}(x, \Gamma_k)$, où

$$\begin{aligned} \mu_{k+1} &= \mu_k + \frac{1}{n+1} (X_{k+1} - \mu_k), k \geq 0, \\ \Gamma_{k+1} &= \Gamma_k + \frac{1}{n+1} \left((X_{k+1} - \mu_k)(X_{k+1} - \mu_k)^T - \Gamma_k \right), \end{aligned} \quad (3.49)$$

On a choisi pour cet algorithme de mettre à jour les paramètres une première fois après 50 itérations puis toutes les 10 itérations.

3.5. APPLICATION

Nous lançons ces algorithmes pour 10000 itérations, ce qui représente déjà un nombre conséquent d'itérations pour des applications au calage d'historique.

Nous allons comparer les résultats produits par chacun de ces quatre algorithmes sur le problème d'inversion bayésienne défini dans le paragraphe 3.5.1 précédent. Pour ce faire, nous introduisons 3 critères :

1. l'estimateur par noyau de la densité (KDE) du posterior sur les figures 3.3, 3.4, 3.5, 3.6, à comparer avec la figure 3.2. Cette dernière représente la KDE du posterior construite à partir d'un échantillon i.i.d de taille 10000 selon la loi a posteriori. Dans la suite, on considère que la fenêtre employée pour l'estimation KDE est choisie en fonction de la taille de l'échantillon.
2. le taux empirique d'acceptation de la transition en figure 3.8,
3. la moyenne des autocorrélations des 2 composantes le long de chaque chaîne en figure 3.9.

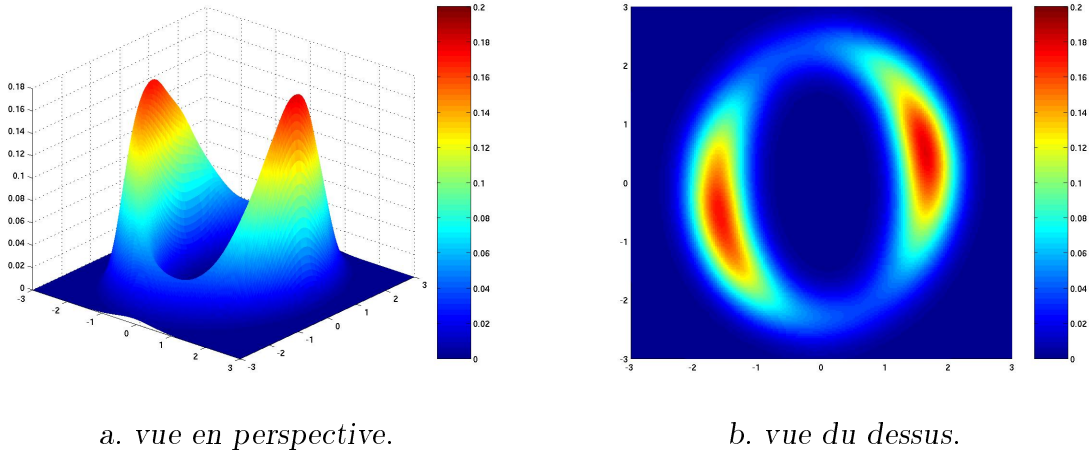


FIG. 3.2 – KDE du posterior construit à partir d'un échantillon i.i.d. de taille 10000.

Sur les figures 3.3 – 3.6, on peut voir que les KDE estimées à partir de chacune des 4 chaînes.

Il est important de remarquer que les algorithmes SIMH et LMH n'arrivent à identifier qu'un seul des 2 modes du posterior (figures 3.4, 3.5), autrement dit, les deux chaînes générées ne balayent pas entièrement le support de π . Cela se traduit par une mauvaise estimation de la loi a posteriori : les KDE respectives ne reproduisent pas les deux modes de π .

En revanche, l'algorithme IMH montre de très bonnes performances en terme de reconstitution de la densité : la KDE obtenue est très satisfaisante (figure 3.3).

Enfin, la chaîne générée par l'algorithme ASIMH donne de très mauvais résultats en terme de KDE (figure 3.6). Bien que le support de π soit bien exploré par la chaîne, on voit apparaître des artefacts dus à la nature non-homogène de l'algorithme. Notamment, le mode que l'on voit apparaître dans le voisinage du point (0.5, 2.5) est certainement dû à la procédure de mise-à-jour. On peut expliquer ce phénomène de la façon suivante : les itérations qui précèdent la première mise à jour des paramètres sont effectuées par une marche aléatoire, son caractère local va induire une estimation

locale des paramètres. Par exemple la variance estimée pour un échantillon très localisé va être très petite, et risque ainsi de créer un artefact. Il semble que les paramètres convergent par la suite (voir figures 3.7).

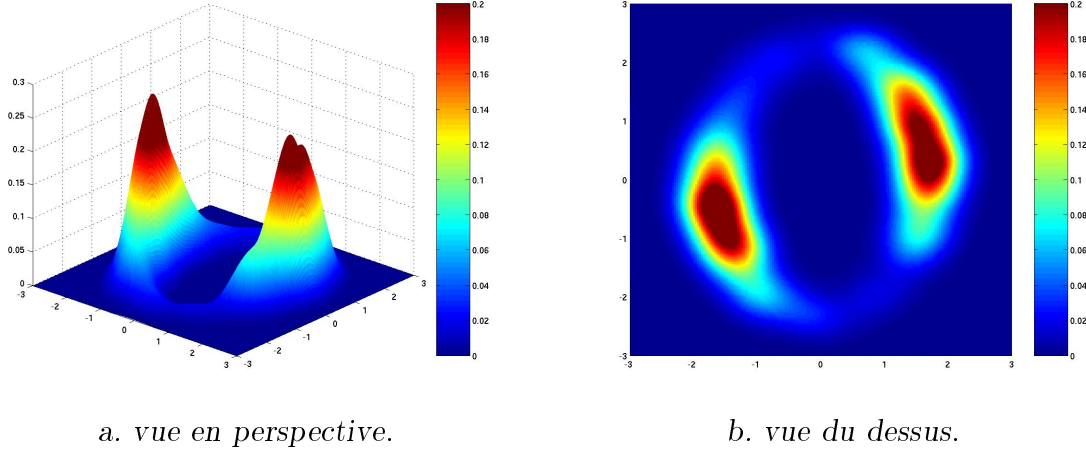


FIG. 3.3 – KDE du posterior construit à partir de la chaîne générée par l'algorithme IMH.

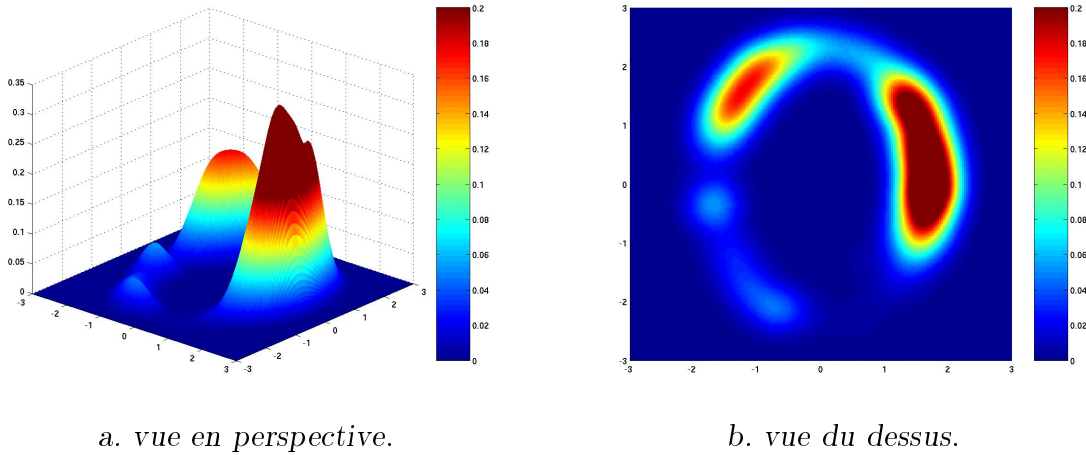


FIG. 3.4 – KDE du posterior construit à partir de la chaîne générée par l'algorithme SIMH.

On représente l'évolution des paramètres de l'ASIMH en figure 3.7.

On peut voir en figure 3.7 l'évolution des paramètres adaptés de l'algorithme ASIMH au cours de la simulation. Les suites (μ_i) et (σ_i^2) représentent respectivement la moyenne et la variance adaptée de la composante i , $(\gamma_{1,2})$ représente la covariance adaptée des deux composantes. Dans un premier temps, on peut remarquer que tous ces paramètres semblent converger, ce qui nous rassure sur la validité de l'algorithme dans cette application (voir théorème 3.4.6). On peut voir que les moyennes et variances semblent converger vers des valeurs satisfaisantes. Le fait que σ_2^2 soit supérieur à σ_1^2 s'accorde avec la représentation du posterior en figure 3.1 où la seconde composante montre une plus importante variabilité que la première. Enfin, la convergence de la covariance

3.5. APPLICATION

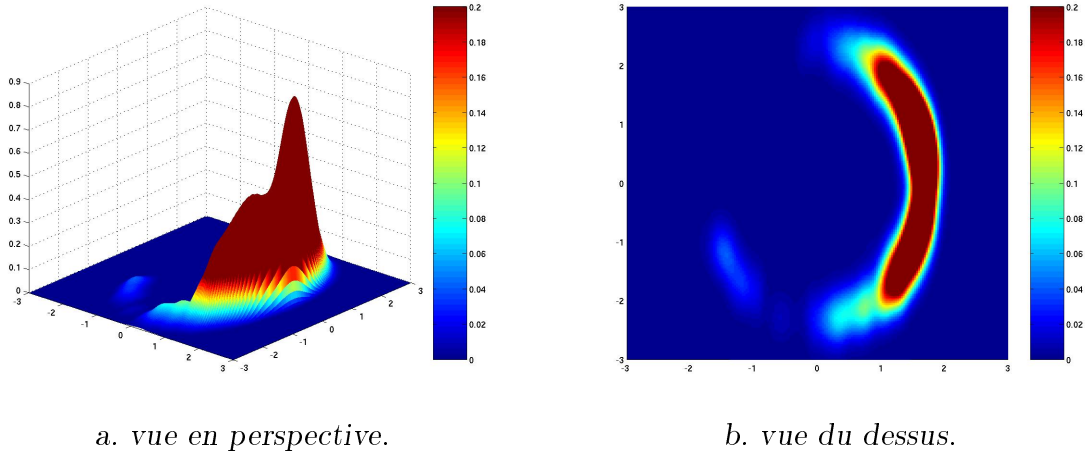


FIG. 3.5 – KDE du posterior construit à partir de la chaîne générée par l'algorithme LMH.

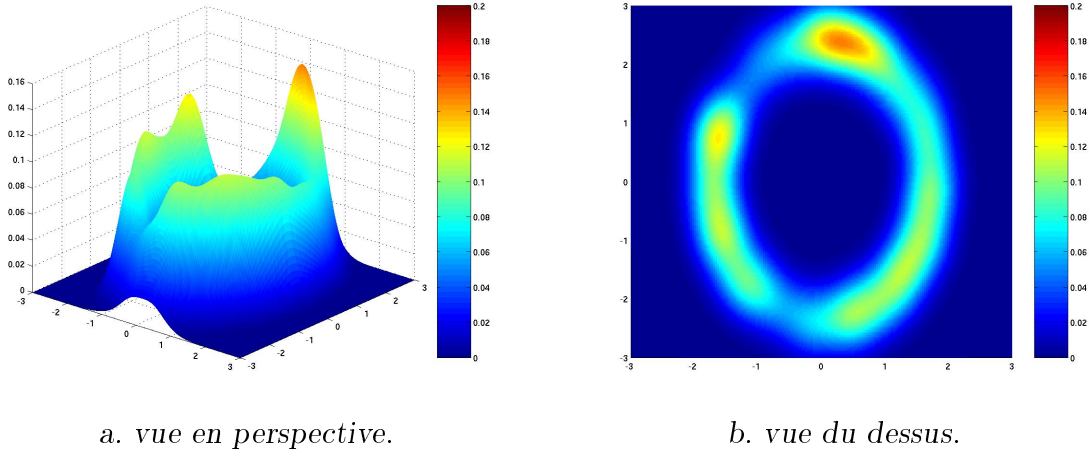


FIG. 3.6 – KDE du posterior construit à partir de la chaîne générée par l'algorithme ASIMH.

empirique vers 0 n'est pas déterminante sur ce problème : on peut voir que les deux composantes présentent des corrélations variables localement.

On peut voir que μ_2 se trouve dans un premier temps aux alentours de 2, avant d'entamer sa convergence vers 0, tandis que μ_1 et σ_1 sont proches de 0. Ce phénomène est en partie responsable des artefacts présentés par la KDE estimée présentée en figure 3.6 : certaines zones de l'ellipse sont très fortement chargées, notamment à l'intersection de l'ellipse support de la loi a posteriori et l'axe $x_1 = 0.5$.

Lorsque l'on regarde les valeurs terminales des paramètres adaptés, il semble qu'ils soient les paramètres d'une gaussienne qui serait très bien adaptée à la simulation de (3.48) par la méthode d'acceptation-rejet, ce qu'illustre le comportement du taux d'acceptation empirique lié à cet algorithme en figure 3.8. Ce comportement disqualifie cet algorithme pour notre domaine d'application.

En ce qui concerne les taux d'acceptation de transition empiriques, les algorithmes IMH et ASIMH montrent les performances les plus mauvaises (figure 3.8), autour de

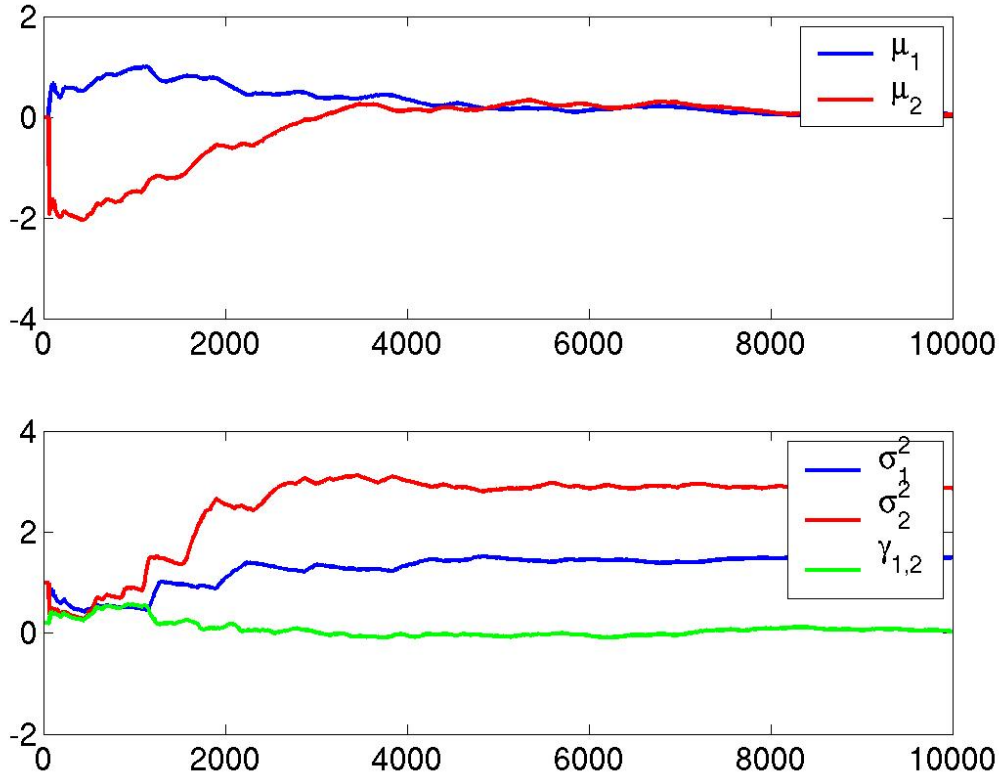


FIG. 3.7 – Evolution des paramètres adaptés au cours de l'algorithme ASIMH.

0.1, ce qui signifie que seulement 10% des transitions sont acceptées, pour ce problème relativement simple. On aurait peut-être obtenu de meilleurs résultats pour l'IMH en choisissant une distribution instrumentale plus adaptée, néanmoins, le choix de la distribution a priori comme loi instrumentale nous semble naturel dans ce contexte.

Au contraire, les algorithmes SIMH et LMH exhibent de bonnes performances au regard de ce critère, particulièrement le LMH avec un taux d'acceptation empirique au-delà de 0.95.

En revanche, les longueurs d'autocorrélation le long des chaînes générées par ces deux algorithmes sont importantes et pénalisent l'estimation de \mathcal{I} , contrairement à l'IMH et l'ASIMH.

Finalement, à la vue de ces différentes figures, il semble que l'on doive choisir entre de bonnes propriétés de mélange, avec des autocorrélations courtes (IMH) mais avec de très faibles taux d'acceptation de transition; et des explorations plus locales avec des transitions très souvent acceptées mais entraînant un mélange lent, autrement dit une décroissance lente des autocorrélations. De plus l'effet d'un accroissement de la dimension, qui n'est pas abordé ici, est bien connu pour empirer les propriétés de la chaîne. En particulier, l'algorithme IMH implémenté ici à titre d'exemple, est très rarement employé dans les problèmes de grande dimension, à moins de connaître très bien la forme de π , afin de construire une distribution instrumentale qui entraînera un

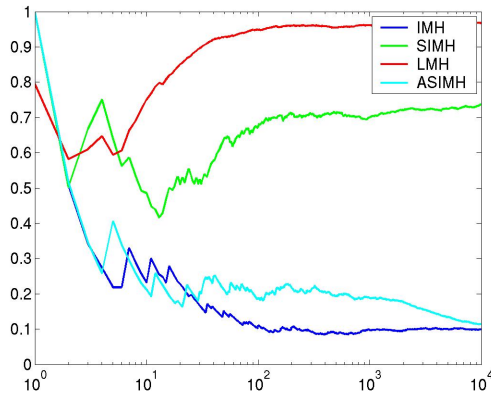


FIG. 3.8 – Taux d'acceptation empirique le long de la chaîne pour chacun des quatre algorithmes, échelle logarithmique.

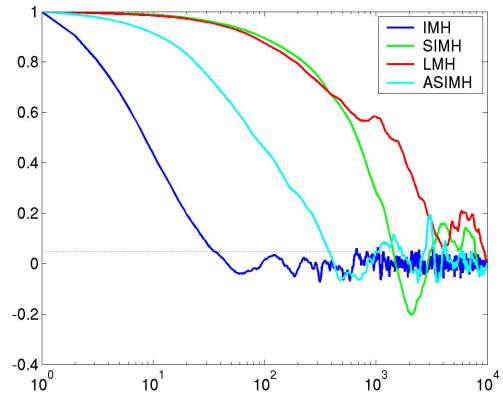


FIG. 3.9 – Autocorrélations le long de la chaîne pour chacun des quatre algorithmes, échelle logarithmique.

taux d'acceptation de transition satisfaisant. Dans le cas des problèmes inverses, cela paraît impossible ; il faudrait alors connaître au préalable les endroits où $\|d - F(X)\|$ est faible, c'est-à-dire que le problème soit déjà résolu.

Pour améliorer les propriétés de mélange de la chaîne sans pour autant accroître son coût informatique, nous introduisons maintenant le principe des chaînes de Markov parallèles en interaction.

3.6 Chaînes en interaction

3.6.1 Introduction

Le principe de faire interagir des chaînes de Markov apparaît pour la première fois dans [Geyer, 1991], sous le nom de Metropolis coupled Markov chain Monte-Carlo (McMCMC), et se développe dans la littérature de physique statistique sous le nom de *parallel tempering* (PT). Ce type d'algorithme a été particulièrement appliqué dans le cadre de simulations physico-chimiques, voir [Earl and Deem, 2005] et sa bibliographie. On le trouve dans la littérature sous différentes appellations : *exchange Monte-Carlo*, *Metropolis coupled-chain*, voir [Iba, 2001] pour une revue.

Le principe de PT est de simuler un nombre $(K + 1)$ de répliques du système d'intérêt par MCMC, chacune à une température différente, au sens de la trempe simulée (section 3.4.8), et d'autoriser les chaînes à échanger de l'information en échangeant leurs états courants. Les systèmes à haute température permettent généralement d'échantillonner de large volumes de l'espace d'état, tandis que ceux à basse température, tout en effectuant un échantillonnage précis localement, peuvent rester bloqués dans un minimum d'énergie local dans le temps d'une simulation sur ordinateur. Le *parallel tempering* réalise une amélioration de l'échantillonnage en autorisant les chaînes à différentes températures à échanger leur configuration. Ainsi, l'inclusion de systèmes à haute température permet aux systèmes à basse température d'accéder à un ensemble représentatif de régions à faible énergie de l'espace d'état.

La simulation de $(K + 1)$ répliques, plutôt qu'une, nécessite de l'ordre de $(K + 1)$ fois le même coût informatique. Cette dépense supplémentaire du PT est une des raisons de la lente adoption de la méthode par les praticiens. En revanche, il est clair qu'une simulation PT est plus que $K + 1$ fois plus efficace qu'une simulation MCMC classique. Cette efficacité vient du fait que l'on permet aux systèmes à basse température d'échantillonner des zones de l'espace d'état qu'ils n'auraient pu atteindre, même en échantillonnant de manière classique pour $(K + 1)$ fois plus d'itérations. Il est également important de noter que l'algorithme PT peut faire un usage très efficace des architectures informatiques en parallèle (*clusters* de processeurs), dont le déploiement s'accélère de nos jours, où les différentes répliques peuvent être lancées en parallèle, contrairement à l'échantillonnage MCMC classique qui est une méthode séquentielle. Un bénéfice supplémentaire offert par le PT est la génération de résultats pour une échelle entière de températures, ce qui peut être d'un grand intérêt pour le praticien, en particulier pour des applications de physique statistique. Le PT est dorénavant largement reconnu comme une méthode numérique utile et puissante.

Plus récemment, des chercheurs de la communauté statistique ont porté leur attention au PT et plus généralement aux chaînes de Markov en interactions. Deux principaux algorithmes ont attiré notre attention. Il s'agit de l'*equi-energy sampler* (EES) de [Kou et al., 2006] et du *population importance resampling MCMC* (PIR) de [Andrieu et al., 2007]. Dans ce dernier article, les auteurs proposent un cadre général pour l'écriture de ces algorithmes ainsi que des résultats de convergence. On trouve également des résultats de convergence de ces algorithmes dans [Atchadé, 2007], néanmoins dans un cadre plus restreint que celui de [Andrieu et al., 2007]. Plus précisément, ces deux algorithmes sont basés sur des approximations de noyaux de Markov non-linéaires, définis dans [Andrieu et al., 2007], et autorisent des échanges d'information avec le passé entier des différentes chaînes. Nous décrivons maintenant le principe de ces méthodes dans le cadre de l'inversion bayésienne. Nous présentons ensuite les algorithmes puis donnons les principaux résultats de convergence.

3.6.2 Principe général

Si l'on cherche à simuler selon la loi (3.1), le principe des méthodes de chaînes en interaction peut être formalisé de la manière suivante. L'idée est de simuler selon la densité :

$$\Pi(X) = \prod_{l=0}^K \pi_l(X), \quad (3.50)$$

où

$$\pi_l(X) \propto e^{-\frac{1}{T_l} \left(\frac{1}{2} \|d - F(X)\|_{C_d^{-1}}^2 - \frac{1}{2} \|X - \mu\|_{\Gamma^{-1}}^2 \right)},$$

et les T_i sont tels que $1 = T_0 < T_1 < \dots < T_K < \infty$. Pour simplifier l'expression de la loi ci-dessus, on peut remarquer qu'elle peut s'exprimer comme une densité de Gibbs, $\pi_l(X) \propto e^{-\frac{1}{T_l} E(X)}$, où $E(X) = \frac{1}{2} \|d - F(X)\|_{C_d^{-1}}^2 + \frac{1}{2} \|X - \mu\|_{\Gamma^{-1}}^2$ représente l'énergie du système. On simule donc $K + 1$ chaînes de loi stationnaire π_l . Dans le schéma classique des méthodes de chaînes en interaction, des *échanges* (*swaps*) sont proposés selon une certaine probabilité entre l'état courant X_n^l de la chaîne l , à température T_l , et un état

3.6. CHAÎNES EN INTERACTION

du passé X_m^{l+1} , $m \leq n$, de la chaîne à la température adjacente supérieure T_{l+1} . Ces échanges sont alors acceptés ou rejetés avec la probabilité :

$$\alpha_{\text{swap}} = \min\left(1, \frac{\pi_l(X_m^{l+1})\pi_{l+1}(X_n^l)}{\pi_l(X_n^l)\pi_{l+1}(X_m^{l+1})}\right) \quad (3.51)$$

$$= \min\left(1, e^{\left(\frac{1}{T_l} - \frac{1}{T_{l+1}}\right)(E(X_n^l) - E(X_m^{l+1}))}\right) \quad (3.52)$$

L'idée est que les distributions π_l à hautes températures sont plus aisées à simuler. En effet, le changement d'échelle opéré par l'emploi de hautes températures va dilater la surface de E , étendant ainsi la surface de ses bassins d'attraction. Une chaîne de Markov de loi stationnaire π_l , moyennant le choix d'un noyau adapté, parcourra ainsi le support de cette loi plus rapidement que les chaînes à basse température. Dans le même temps, les échanges vont permettre d'apporter l'information vers les chaînes à basse température au comportement plus local.

Ces algorithmes posent néanmoins un certain nombre de difficultés pratiques, portant sur le choix des paramètres, que nous abordons en section 3.6.5.

3.6.3 Description des algorithmes

Avant la description des algorithmes, nous introduisons la notion de noyau de Markov non-linéaire. Elle nous permettra de présenter les 3 algorithmes cités plus haut dans un même cadre.

Noyaux de Markov non-linéaires

Les méthodes MCMC standards reposent sur l'emploi de noyaux de Markov de la forme $K : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{X})$. Ces noyaux sont des opérateurs linéaires sur $\mathcal{P}(\mathcal{X})$ dans le sens où $\mu(dy) = \int_{\mathcal{X}} \nu(dx)K(x, dy)$, où $\mu, \nu \in \mathcal{P}(\mathcal{X})$. Un noyau de Markov non-linéaire $K_\mu : \mathcal{P}(\mathcal{X}) \times \mathcal{X} \rightarrow \mathcal{P}(\mathcal{X})$ est un opérateur non-linéaire sur $\mathcal{P}(\mathcal{X})$. De tels noyaux K_μ peuvent être construits pour améliorer les propriétés de mélange des noyaux linéaires. On définit ainsi, pour $x \in \mathcal{X}$, $A \in \mathcal{B}(\mathcal{X})$:

$$K_\mu(x, A) = (1 - \theta)K(x, A) + \theta\Phi(\mu)(A), \quad (3.53)$$

où K est un noyau de Markov de distribution invariante π , $\theta \in]0, 1[$, $\Phi : \mathcal{P}(E) \rightarrow \mathcal{P}(E)$ est un opérateur de sélection/mutation, tel que $\Phi(\mu) := \mu(gK)/\mu(g)(dy)$ et g est une fonction. La fonction g est choisie bornée et mesurable, et telle que $\Phi(\pi) = \pi$. La simulation selon K_π a de très bonnes propriétés théoriques : elle permet des régénérations selon π et le noyau K_π est fortement uniformément ergodique (cf. [Andrieu et al., 2007] et sa bibliographie). Cependant, il est impossible de la réaliser : elle suppose le problème résolu. Les auteurs proposent alors une approximation basée sur le passé de la chaîne.

Typiquement, μ est remplacée par la mesure empirique $\hat{\mu}_n^X = \frac{1}{n+1} \sum_{i=0}^n \delta(X_i)$, où les X_i sont les n premières réalisations de la chaîne.

Ainsi, on approxime $\Phi(\mu)(A)$ par :

$$\Phi(\hat{\mu}_n^X)(A) = \frac{\sum_{i=0}^n g(X_i)K(X_i, A)}{\sum_{i=0}^n g(X_i)} \quad (3.54)$$

Simuler selon ce noyau présente peu d'intérêt lorsque l'on simule une seule chaîne, elle interagira alors uniquement avec son propre passé. En revanche dans le cas d'une population de chaînes à différentes températures, où l'approximation (3.54) est construite à partir d'une chaîne à température supérieure, cela peut s'avérer très intéressant.

L'algorithme

On définit d'abord la famille $\{\pi^{(l)}, l = 0..K\}$ de distributions selon lesquelles nous souhaitons simuler :

$$\pi^{(l)}(X) \propto e^{-E_l(X)}, \quad (3.55)$$

où $E_l(X) = \frac{E(X)}{T_l}$, avec $E(X)$ l'énergie du système, et T_l la température à laquelle il est considéré. Les T_l sont définies comme suit : $T_0 = 1 < T_1 < \dots < T_K < +\infty$, et donc $\pi^{(0)} = \mathbb{P}(X|d)$. On notera $x = (x^{(0)}, \dots, x^{(K)}) \in \mathcal{X}^{K+1}$ l'état courant de la chaîne. Pour $l = 0, \dots, K-1$, on définit la fonction d'importance :

$$r^{(l)}(x^{(l)}) = e^{-(E_l(x^{(l)}) - E_{l+1}(x^{(l)}))} \quad (3.56)$$

La méthode peut être formalisée en introduisant le noyau inhomogène \mathbf{K}_n , étant donné l'ensemble des états passés $x_{0:n-1} = (x_0, \dots, x_{n-1})$ et pour $A_0 \times \dots \times A_K \in \mathcal{B}(\mathcal{X}^{K+1})$:

$$\mathbf{K}_n(x_{0:n-1}; A_0 \times \dots \times A_K) = K^{(K)}(x^{(K)}, A_K) \prod_{l=0}^{K-1} \mathbf{K}_{x_{0:n-1}}^{(l)}(x^{(l)}, A_l), \quad (3.57)$$

où pour $x_{0:n-1}^{(l+1)} \in \mathcal{X}^n$, $x^{(l)} \in \mathcal{X}$ et $A \in \mathcal{B}(\mathcal{X})$:

$$\mathbf{K}_{x_{0:n-1}}^{(l)}(x^{(l)}; A) = \theta K^{(l)}(x^{(l)}, A) + (1 - \theta) \int_{\mathcal{X}} \nu_{x_{0:n-1}}^{(l)}(x^{(l)}, dy) T^{(l)}(y, x^{(l)}; A), \quad (3.58)$$

et

$$\nu_{x_{0:n-1}}^{(l)}(x^{(l)}, dy) = \frac{\sum_{i=0}^{n-1} \omega_{n,i}^{(l)}(x^{(l)}, x_i^{(l+1)}) \delta_{x_i^{(l+1)}}(dy)}{\sum_{i=0}^{n-1} \omega_{n,i}^{(l)}(x^{(l)}, x_i^{(l+1)})} \quad (3.59)$$

et dans les algorithmes ici considérés, $T^{(l)}$ prendra la forme suivante :

$$T^{(l)}(y, x^{(l)}; A) = \min \left(1, \frac{r^{(l)}(y)}{r^{(l)}(x^{(l)})} \right) \mathbb{1}_A(y) + \left(1 - \min \left(1, \frac{r^{(l)}(y)}{r^{(l)}(x^{(l)})} \right) \right) \mathbb{1}_A(x^{(l)}) \quad (3.60)$$

Le noyau $\mathbf{K}_{x_{0:n-1}}^{(l)}(x^{(l)}; A)$ est ainsi un noyau non-linéaire analogue à (3.53). En d'autres termes, à l'itération n , à la température T_l , avec probabilité θ , une transition MH classique sera réalisée. Sinon, avec probabilité $(1-\theta)$, un échange sera proposé. Il consiste à choisir un état y parmi $x_{0:n-1}^{(l+1)}$ selon les poids $\omega_{n,i}^{(l)}(x^{(l)}, x_i^{(l+1)})$. Cet échange est alors

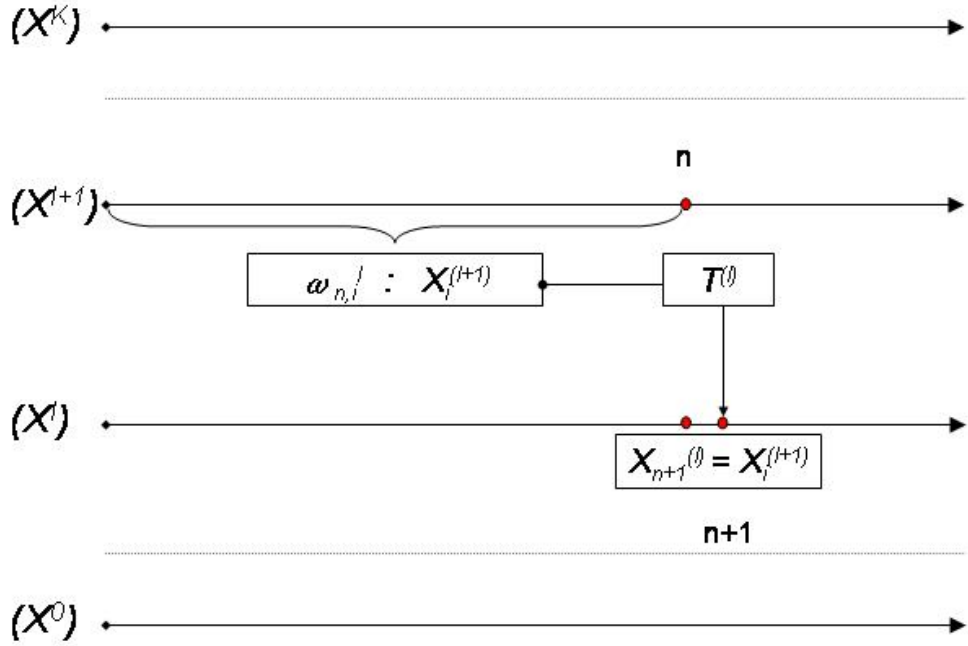


FIG. 3.10 – Représentation schématique d'un échange.

accepté avec la probabilité $\min\left(1, \frac{r^{(l)}(y)}{r^{(l)}(x^{(l)})}\right)$. Le nouvel état est choisi selon $\nu_{x_{0:n-1}}^{(l)}$, ce qui peut être vu comme une étape d'échantillonnage d'importance : on simule selon $\pi^{(l)}$ avec la loi instrumentale $\pi^{(l+1)}$. Le principe de l'échange ou interaction entre deux chaînes est schématisé en figure 3.10. Nous y représentons l'occurrence d'un échange entre les chaînes X^{l+1} et X^l à l'itération n . Les trois algorithmes (PT, EES, PIR), considérés ici peuvent être écrits dans ce cadre et ne diffèrent que dans la formulation des poids $\omega_{n,i}^{(l)}$, pour $y, z \in \mathcal{X}$:

- pour l'algorithme PT, on a :

$$\omega_{n,i}^{(l)}(y, z) = \mathbb{1}_{i=n-1}, \quad (3.61)$$

autrement dit, il est possible uniquement d'aller vers l'état courant de la chaîne à la température adjacente supérieure,

- pour l'algorithme PIR, les poids ω prennent la forme suivante :

$$\omega_{n,i}^{(l)}(y, z) = r^{(l)}(z), \quad (3.62)$$

c'est-à-dire que l'on obtient le nouvel état en ré-échantillonnant $x_{0:n-1}^{(l+1)}$ avec les poids ω . On remarquera que cet algorithme diffère de l'algorithme (NL3) de [Andrieu et al., 2007], dénommé (IR-McMC) dans [Atchadé, 2007], en ce que le

nouvel état de la chaîne à température l sera celui tiré en ré-échantillonnant ; dans l'algorithme sus-cité cet état sert de point de départ à partir duquel une nouvelle transition est proposée. Cette nouvelle transition pouvant être rejetée, il nous semble intéressant de supprimer cette dernière étape afin d'augmenter la probabilité de réaliser des échanges.

- pour l'algorithme EES, étant donnée une suite de niveaux d'énergie $A_0 < A_1 < \dots < A_K < A_{K+1} = \infty$, on définit la partition : $\mathcal{X} = \bigcup_{l=0}^K \mathcal{X}_l$ d'anneaux d'énergie : $\mathcal{X}_l = \{x \in \mathcal{X} : A_l < E(x) < A_{l+1}\}$, et la fonction $I(x) = l$ si $x \in \mathcal{X}_l$. Les ω prennent la forme suivante :

$$\omega_{n,i}^{(l)}(y, z) = \mathbb{1}_{\mathcal{X}_{I(y)}}(z). \quad (3.63)$$

Autrement dit, le nouvel état sera choisi uniformément parmi les états $x_{0:n-1}^{(l+1)}$ de la chaîne à la température T_{l+1} qui sont dans le même anneau d'énergie que l'état courant $x^{(l)}$.

L'idée principale sur laquelle repose les deux derniers algorithmes est que le noyau (3.58) va converger vers le noyau limite suivant :

$$\mathbf{K}_{x_{0:n-1}^{(l+1)}}^{(l)}(x^{(l)}; A) = \theta P^{(l)}(x^{(l)}, A) + (1 - \theta) R^{(l)}(x^{(l)}, A), \quad (3.64)$$

où $R^{(l)}$ est le noyau d'une chaîne de type MH, dont la proposition de transition est donné par $Q_{PIR}^{(l)}(x, dy) = \pi^{(l)}(dy)$ dans le cas de l'algorithme PIR, et par $Q_{EES}^{(l)}(x, dy) \propto \pi^{(l+1)}(y) \mathbb{1}_{\mathcal{X}_{I(x)}}(y) \lambda(dy)$ dans le cas de l'EES. L'approximation de (3.64) à temps fini va nous aider à échantillonner selon $\pi^{(0)}$, en particulier en balayant l'espace d'état de manière plus efficace.

Finalement, il est important de noter que pour les 3 algorithmes, on peut employer les échantillons générés par l'ensemble des chaînes, en leur associant les poids d'importance suivant :

$$\eta^{(l)}(x) = e^{-(E_0(x) - E_l(x))}, \quad (3.65)$$

dans le but d'estimer $\mathcal{I} = E_{\pi_0}[h(X)]$, pour un certain h . Ainsi l'estimateur $\widehat{\mathcal{I}}$, après n itérations de l'algorithme, prendra la forme suivante :

$$\widehat{S_n(h)} = \sum_{l=0}^K \frac{\sum_{i=0}^n \eta^{(l)}(x_i^{(l)}) h(x_i^{(l)})}{\sum_{i=0}^n \eta^{(l)}(x_i^{(l)})}. \quad (3.66)$$

On trouve dans [Kou et al., 2006], une justification numérique de l'emploi de l'échantillon entier pondéré qui fournira un meilleur estimateur que si l'on considère uniquement les états de la chaîne à la température minimale $x_{0:n}^{(0)}$.

Nous introduisons maintenant quelques résultats de convergence pour ces algorithmes.

3.6.4 Résultats de convergence

Nous détaillons ici les conditions nécessaires et les résultats de convergence pour chacun des deux algorithmes PIR et EES, dont les démonstrations peuvent être trouvées dans [Atchadé, 2007]. Des résultats de convergence dans le cadre plus général des noyaux de Markov non-linéaires peuvent être trouvés dans [Andrieu et al., 2007].

L'algorithme PIR

Les conditions permettant d'établir une loi des grands nombres et un théorème de la limite centrale pour l'algorithme PIR sont les suivantes :

- (A-1) Pour $l = 0, \dots, K-1$, la fonction $r^{(l)}$, définie en (3.56), est telle qu'il existe une constante $\varepsilon > 0$ tel que $r^{(l)}(x) > \varepsilon$ pour tout $x \in \mathcal{X}$.
- (A-2) Il existe une fonction mesurable $V : \mathcal{X} \rightarrow [1, \infty[$, une mesure ϕ , telle que $K^{(K)}$, qui est le noyau de Markov linéaire correspondant à la plus haute température, est ϕ -irréductible, apériodique et

$$\|K^{(K)^n} - \pi^{(K)}\|_V = O(\rho^n),$$

où $\|T\|_V = \sup_{f \in \mathcal{L}_V, \|f\|_V \leq 1} \|Tf\|_V$, $\mathcal{L}_V = \{f : \mathcal{X} \rightarrow \mathbb{R} : \|f\|_V < \infty\}$, avec les notations de la section 3.4.7, et pour $\rho \in]0, 1[$.

- (A-3) Pour tout $\beta \in [0, 1]$,

$$\sup_{n > 0} \|K^{(l)^n} V^\beta\|_{V^\beta} < \infty, \quad l = 0, \dots, K-1,$$

où $K^{(l)}$ est donné en (3.58).

La première condition est une condition de stabilité des algorithmes d'échantillonnage d'importance. La seconde signifie que la chaîne à la température maximale doit converger géométriquement et la troisième est une condition de stabilité aisément remplie dans notre cas. On en déduit le théorème de convergence suivant.

Théorème 3.6.1 *On suppose remplies les conditions (A1-3) et $l \in \{0, \dots, K\}$ fixé.*

- i) Soit $f \in \mathcal{L}_{V^\beta}$, $\beta \in [0, 1/4[$, alors

$$\frac{1}{n} \sum_{i=1}^n f(X_i^l) \rightarrow \pi^{(l)}(f), \quad p.s. \quad (3.67)$$

- ii) Pour $\beta \in [0, 1/8[$, soit $f \in \mathcal{L}_{V^\beta}$ telle que $\pi^{(l)}(f) = 0$. On définit

$$\sigma_l^2(f) = \pi^{(l)}(f^2) + 2 \sum_{i=1}^{\infty} \theta^i \int_{\mathcal{X}} f(x) K^{(l)^i} f(x) \pi^{(l)}(dx),$$

et l'on suppose que $\sigma_l^2(f) > 0$. Il existe alors une suite aléatoire $\{\pi_n^{(l)}(f)\}$, telle que $\pi_n^{(l)}(f) \rightarrow \pi^{(l)}(f)$ presque sûrement, telle que :

$$\frac{1}{\sqrt{n\sigma_l(f)}} \sum_{i=1}^n [f(X_i^{(l)}) - \pi_i^{(l)}(f)] \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1). \quad (3.68)$$

L'algorithme EES

Une condition qui permet d'établir les théorèmes limites pour l'EES est :

(B) Pour $l = 0, \dots, K-1$, il existe un ensemble $C_l \subset \mathcal{X}$, une mesure de probabilité ϕ_l telle que $\phi_l(C_l) > 0$ et des constantes $\lambda_l \in (0, 1)$, $b_l \in [0, \infty)$, $\varepsilon \in (0, 1]$, telles que pour tout $x \in \mathcal{X}$ et $A \in \mathcal{B}(\mathcal{X})$,

$$K^{(l)}(x, A) \geq \varepsilon_l \phi_l(A) \mathbf{1}_{C_l}(x), \quad (3.69)$$

et

$$K^{(l)}(x, A)V(x) \leq \lambda_l V(x) + b_l \mathbf{1}_{C_l}(x), \quad (3.70)$$

où $V(x) = c \exp(\kappa E(x)) \geq 1$, avec $c \in (0, \infty)$, et $\kappa \in (0, 1)$. De plus κ satisfait :

$$0 < \theta \lambda_l + (1 - \theta) \left(1 + \frac{\kappa}{\frac{1}{T_l} - \frac{1}{T_{l+1}} - \kappa} \right) < 1, \quad l = 0, \dots, K-1 \quad (3.71)$$

On peut vérifier (B) pour de nombreux exemples pratiques. Par exemple, si chaque $K^{(l)}$ est une marche aléatoire à accroissements symétriques, alors (3.69) et (3.70) sont vérifiées, cf. [Robert and Casella, 2004]. Dans la plupart des cas, il est toujours possible de choisir $\kappa \in (0, 1)$ assez petit pour que (3.71) soit vérifiée. Les résultats de convergence prennent une forme semblable à ceux du PIR :

Théorème 3.6.2 *On suppose (B) et $l \in \{0, \dots, K\}$ fixé.*

i) Soit $f \in \mathcal{L}_{V^\beta}$, $\beta \in [0, 1/2[$, alors

$$\frac{1}{n} \sum_{i=1}^n f(X_i^{(l)}) \rightarrow \pi^{(l)}(f), \quad p.s. \quad (3.72)$$

ii) Pour $\beta \in [0, 1/4[$, soit $f \in \mathcal{L}_{V^\beta}$ telle que $\pi^{(l)}(f) = 0$. On définit

$$\sigma_l^2(f) = \pi^{(l)}(f^2) + 2 \sum_{i=1}^{\infty} \theta^i \int_{\mathcal{X}} f(x) \mathbf{K}^{(l)^i} f(x) \pi^{(l)}(dx),$$

où $\mathbf{K}^{(l)}$ est donné en (3.58). On suppose que $\sigma_l^2(f) > 0$. Il existe alors une suite aléatoire $\{\pi_n^{(l)}(f)\}$, telle que $\pi_n^{(l)}(f) \rightarrow \pi^{(l)}(f)$ presque sûrement, telle que :

$$\frac{1}{\sqrt{n} \sigma_l(f)} \sum_{i=1}^n [f(X_i^{(l)}) - \pi_i^{(l)}(f)] \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1). \quad (3.73)$$

Pour les deux algorithmes, la validité de l'estimateur (3.66) découle de ces résultats.

3.6.5 Considérations pratiques

Les méthodes présentées ici sont très récentes et peu de conseils pratiques sont disponibles à leur sujet. A partir des données de la littérature et de notre propre expérience, nous allons discuter des 4 points suivants, dans le cadre de l'inversion bayésienne :

1. la construction de l'échelle de température,
2. le choix du noyau en fonction de la température,
3. le nombre de chaînes,
4. la probabilité de proposition d'un échange.

La première difficulté liée à ces algorithmes est le choix du nombre et de la répartition des T_l . Il faut en pratique assurer le chevauchement des densités des énergies non-normalisées ($\frac{E(X)}{T_l}$), ce qui permettra aux échanges d'être acceptés, le cas échéant. Cependant, dans le cas des problèmes inverses, le chevauchement est naturel puisque l'énergie minimale $\min_{x \in \mathcal{X}}(E(x))$ est une valeur fixe : la quantité ($\frac{E(X)}{T_l}$) prendra ainsi des valeurs plus faibles pour les hautes températures. La surface de chevauchement entre les densités de ($\frac{E(X)}{T_l}$) doit cependant être assez importante, afin d'accepter les échanges.

Pour ce qui est de la répartition des β_i , une répartition géométrique est généralement préconisée. Par exemple, si l'on fixe $T_0 = 1$ et T_{K+1} , on choisira $\beta_i = \prod_{j=1}^{K+1} \sqrt{\frac{T_{K+1}}{T_0}}$. Nous apportons à la simulations selon $\pi = \pi_0$, qui est la loi cible, les informations glanées à plus haute température via les échanges.

Comme dit plus haut, l'idée de l'application de ces chaînes est l'amélioration du mélange de la chaîne. Il faut alors choisir des noyaux qui rendent cette propriété effective. Aux plus hautes températures, des transitions *globales* sont préconisées : $\pi^{(K)}$ étant une version dilatée de $\pi = \pi^{(0)}$, le paysage énergétique correspondant est aplati et des transitions larges peuvent permettre de se déplacer aisément d'un bassin d'attraction à l'autre. Nous proposons alors de choisir un échantillonneur séquentiel indépendant qui affecte un grand nombre de composantes, voire toutes, de X , ou bien une marche aléatoire à accroissements symétriques avec de larges pas.

Aux basses températures, en revanche, il convient d'employer des échantillonneurs *locaux*, puisque, dans le cadre de l'inversion bayésienne, la non-linéarité de l'opérateur direct entraîne une forte complexité des distributions cibles. Une marche aléatoire à accroissements symétriques ou l'échantillonneur de Langevin, qui ont de bonnes propriétés d'exploration locales, sont ainsi préconisés, sinon, les transitions des chaînes risquent d'être très souvent rejetées.

Entre les deux températures extrêmes, il est nécessaire de choisir des noyaux dont le mélange s'améliore lorsque la température augmente tout en gardant des taux d'acceptation de transition et d'échange satisfaisants. A ce titre, il peut être intéressant d'employer des échantillonneurs auto-régressifs (voir section 3.4.5) dont les paramètres seraient fonction de la température.

D'autre part, ces considérations à propos des noyaux sont directement liées aux nombres de chaînes à employer. Il est notamment important de ne pas utiliser un nombre trop élevé de chaînes. En effet, l'emploi d'un trop grand nombre de chaînes va ralentir la descente de l'information des températures les plus hautes vers les plus basses. Pour autant, le nombre de chaînes doit être assez élevé afin d'autoriser les échanges. Dans la littérature relative au Parallel Tempering, on trouve comme ordre de grandeur du nombre de chaînes à employer \sqrt{d} où d est la dimension de \mathcal{X} . D'après notre expérience, avec les algorithmes EES et PIR, il semble qu'un nombre bien moins important suffise.

Enfin, en ce qui concerne le choix de la probabilité de proposer un échange, dans

[Kou et al., 2006] les auteurs proposent de le fixer dans l'intervalle $[0.2, 0.3]$, sans plus d'explication. À la lumière de notre expérience, il nous semble que ce paramètre soit particulièrement sensible. Une valeur trop faible entraînera peu d'échanges et les bonnes propriétés de mélange de l'algorithme s'en verront atténuées ; une valeur trop élevée impliquera de nombreux retours vers des régions déjà explorées dont les éléments se verront attribués des poids ω élevés. Il s'agit donc de trouver un juste milieu, ce qui dépend largement du problème considéré.

En conclusion, on peut dire à propos de ces 4 points qu'il faut réaliser un bon compromis afin d'autoriser un échange d'information efficace, tout en réalisant une bonne exploration locale à basse température et globale à haute température. Ces choix sont à effectuer relativement au problème considéré.

3.6.6 Application

Nous comparons ici les résultats produits par ces trois algorithmes sur le problème défini en section 3.5.1. Nous considérons 4 températures différentes ($T_3 = 50$, $T_2 = 13.5722$, $T_1 = 3.6841$, $T_0 = 1$) et nous fixons la probabilité d'échange θ à 0.2. À chaque température, nous employons un algorithme SIMH, dont la variance du pas varie selon la température, plus précisément : 0.1 à T_0 et $0.1\sqrt{T_l}$ pour $l = 1, 2, 3$.

Nous n'utilisons pas ici la totalité des critères définis en section 3.5.2. En effet, le taux d'acceptation empirique de chaque chaîne au sein de chacun des 3 algorithmes se comportera de la même manière que celui d'un SIMH classique. Nous rappelons toutefois les deux critères définis en 3.5.2 que nous étudions ici :

1. la moyenne des autocorrélations le long de chacune des chaînes parallèles en figures 3.11, 3.12 and 3.13,
2. la KDE du posterior en figures 3.14, 3.15, et 3.16 à comparer avec 3.2. Ici aussi, nous employons une fenêtre fonction de la taille de l'échantillon. Cet estimateur est construit en considérant l'échantillon total, c'est-à-dire contenant les états des 4 chaînes, en leur associant les poids (3.65).

En étudiant les figures 3.11, 3.12 et 3.13, on peut voir que ces méthodes présentent l'avantage majeur de réduire les autocorrélations, par rapport au MCMC simple. En effet, bien qu'à chaque température nous employions un échantillonneur *local*, la longueur des autocorrélations est très courte. Ce phénomène s'explique par l'emploi d'échanges entre les chaînes. En effet, un échange correspond à une *régénération* de la chaîne, un nouvel état qui sera indépendant de l'état précédent, et son occurrence le long de chaque chaîne réduit de fait les autocorrélations. Les performances des trois algorithmes sont sensiblement les mêmes.

Sur les figures 3.14, 3.15 et 3.16, on représente respectivement les KDE des échantillons résultants des algorithmes PT, EES et PIR. Elles rendent bien compte de l'efficacité de ces méthodes pour échantillonner selon cette loi multimodale à support complexe. Les densités estimées reproduisent de manière satisfaisante les deux modes.

On a montré sur cette application l'avantage d'employer des méthodes de chaînes de Markov en interaction : elles permettent d'accélérer le balayage du support de la loi cible

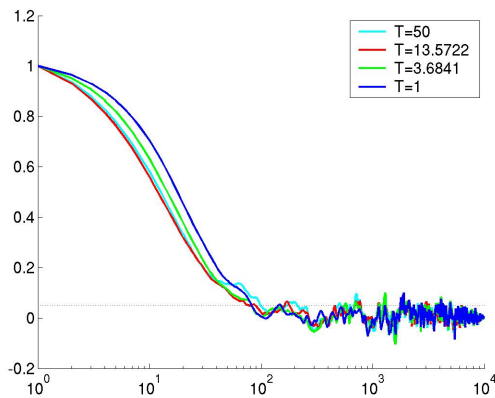


FIG. 3.11 – Autocorrélations le long de la chaîne pour l'algorithme PT, échelle logarithmique.

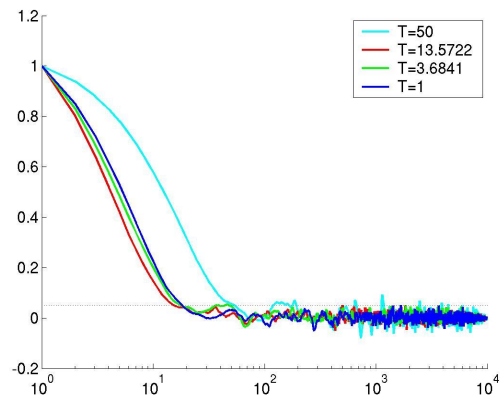


FIG. 3.12 – Autocorrélations le long de la chaîne pour l'algorithme EES, échelle logarithmique.

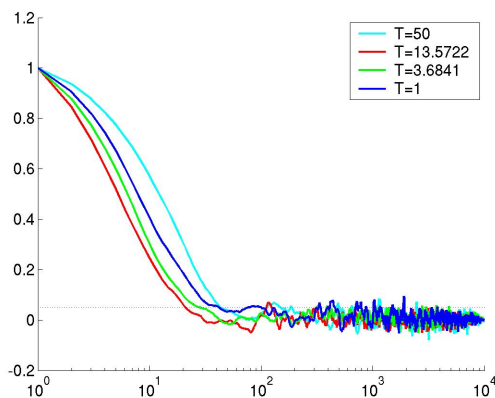


FIG. 3.13 – Autocorrélations le long de la chaîne pour l'algorithme PIR, échelle logarithmique.

tout en réduisant les autocorrélations par rapport aux algorithmes MH classiques. Ceci est particulièrement intéressant pour l'estimation de \mathcal{I} . On peut également remarquer que ces performances sont réalisées sans accroître le coût informatique : elles sont particulièrement adaptées au calcul parallèle. En vérité, à nombre d'itérations n fixé, l'algorithme parallélisé est même plus rapide puisque qu'il nécessite en moyenne $(1 - \theta) \times n$ calculs de l'opérateur direct au total : il n'est pas effectué lorsque un échange est proposé. Ces méthodes semblent ainsi parfaitement adaptées à la résolution du problème inverse dans le cadre bayésien c'est-à-dire à la simulation selon (3.1). Nous l'appliquons au problème du calage d'historique dans le chapitre suivant.

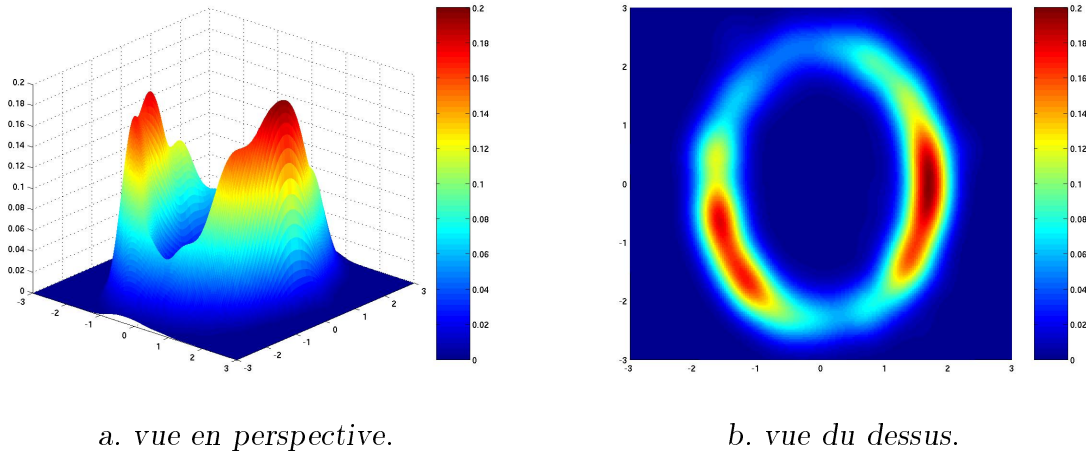


FIG. 3.14 – KDE du posterior construit à partir des résultats de l'algorithme PT.

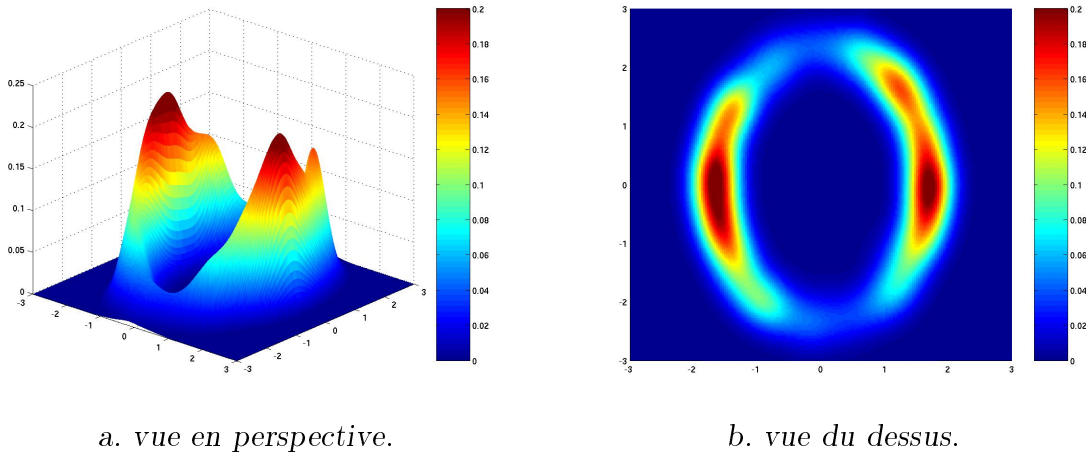


FIG. 3.15 – KDE du posterior construit à partir des résultats de l'algorithme EES.

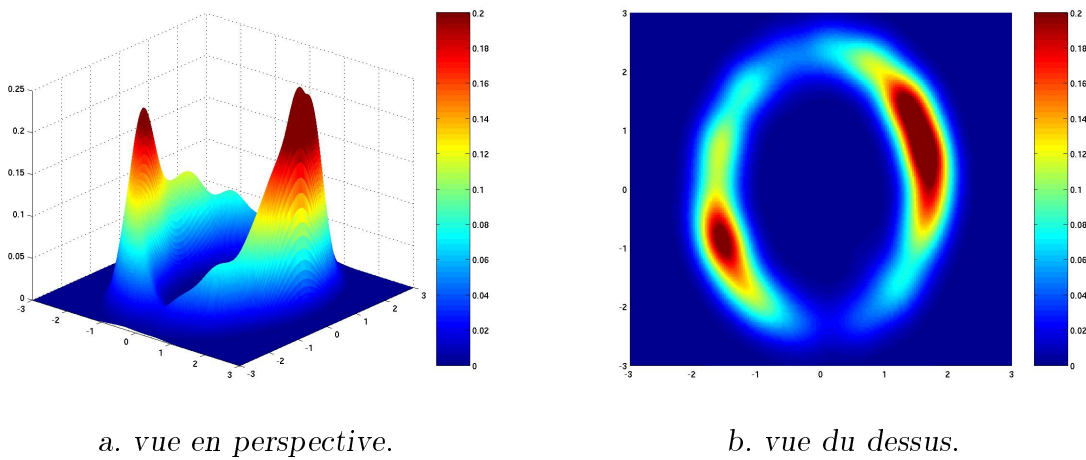


FIG. 3.16 – KDE du posterior construit à partir des résultats de l'algorithme PIR.

Chapitre 4

Applications

Ce chapitre est constitué des 3 articles principaux rédigés durant la thèse.

Le premier s'intitule *Integrating production data under uncertainty by parallel interacting Markov chains on a reduced dimensional space* et paraîtra prochainement dans la revue *Computational Geosciences*. Il reprend la partie applicative, ainsi que l'exposé des méthodes de chaînes de Markov en interaction du chapitre 3 et présente une application de notre méthodologie à un cas synthétique de conditionnement du champ de perméabilité à des données de water cut.

Le second a pour titre *History matching of truncated Gaussian models by parallel interacting Markov chains on a reduced dimensional space*. Il a été publié dans les actes du congrès ECMOR XI, 8 - 11 September 2008 in Bergen, Norway. Il présente une application de notre méthodologie à un cas synthétique de conditionnement d'un modèle à deux faciès à des données de water cut. On y expose également une comparaison avec les résultats fournis par une méthode MCMC classique.

Le troisième article a été réalisé en collaboration avec l'équipe COUGAR de l'IFP. Le logiciel COUGAR est un programme industriel permettant de réaliser des analyses de sensibilité et d'incertitudes sur différents paramètres du réservoir. Nous avons appliqué des méthodes d'analyse de sensibilité en prenant en compte les composantes de la décomposition de Karhunen-Loève du champ de perméabilité. Il s'agit donc d'une approche différente de celle développée dans le manuscrit. Cet article s'intitule *Reservoir forecasting under uncertainty*. Nous en présentons une version préliminaire. Il sera par la suite soumis pour publication dans la revue *Computational Geosciences*.

4.1 Integrating production data under uncertainty by parallel interacting Markov chains on a reduced dimensional space

Integrating production data under uncertainty by parallel interacting Markov chains on a reduced dimensional space

Thomas Romary

Received: date / Accepted: date

Abstract In oil industry and subsurface hydrology, geostatistical models are often used to represent the porosity or the permeability field. In history matching of geostatistical reservoir model, we attempt to find multiple realizations that are conditional to dynamic data and representative of the model uncertainty space. A relevant way to simulate the conditioned realizations is by generating Monte Carlo Markov Chains (MCMC). The huge dimension (number of parameters) of the model and the computational cost of each iteration are two important pitfalls for the use of MCMC. In practice, we have to stop the chain far before it has browsed the whole support of the posterior probability density function. Furthermore, as the relationship between the production data and the random field is highly non-linear, the posterior can be strongly multi-modal and the chain may stay stuck in one of the modes. In this work, we propose a methodology to enhance the sampling properties of classical single MCMC in history matching. We first show how to reduce the dimension of the problem by using a truncated Karhunen-Loève expansion of the random field of interest and assess the number of components to be kept. Then we show how we can improve the mixing properties of MCMC, without increasing the global computational cost, by using parallel interacting Markov Chains. Finally, we show the encouraging results obtained when applying the method to a synthetic history matching case.

Keywords Reservoir characterization · Inverse problem · Karhunen-Loève expansion · Interacting Markov chains · History Matching

1 Introduction

Conditioning the reservoir petrophysical properties, *e.g.* permeability or porosity, to production data, such as cumulative oil production, water cut, is a most challenging task. It consists in solving an ill-posed inverse problem: given a prior knowledge on the random field representing the petrophysical properties of the reservoir, in terms of parameters of a geostatistical model, we aim to find multiple realizations of this model that will exhibit the same dynamical behaviour as the one observed. In other words, we want to sample from the posterior distribution defined in the Bayesian inversion framework. The dynamical behaviour of a given field is computed by a fluid-flow simulator.

A unique realization of a conditioned field can be found using optimization techniques to find the minimum of an objective function. Some parameterization methods can simplify the problem by reducing the dimension of the space where the optimization is performed, see *e.g.* [14], [5]. However, optimization methods are not relevant to cope with uncertainty in reservoir characterization. Although multiple optimizations can be performed and hence multiple conditioned realizations can be found by this technique, the uncertainties cannot be assessed clearly. Kitanidis [16] and Oliver [21] proposed a method to characterize uncertainty with optimization procedures, by perturbing the dynamical data. This method implies to solve a huge number of optimization problems to sample correctly the posterior. Moreover, their method is known to work well with linear or

T. Romary
Reservoir engineering dept., IFP,
1 & 4, avenue de Bois-Preau
92852 Rueil-Malmaison Cedex - FRANCE
Tel.: +33-1-47526000
Fax: +33-1-47525617
E-mail: thomas.romary@ifp.fr

quasi-linear operators. The fluid-flow models are often highly non-linear however.

The Bayesian formulation of inverse problem, see [28], shows that this problem can be viewed as a classical statistical inference problem, where we want to sample independent realizations from the posterior distribution, known up to a constant; the posterior being defined from the prior knowledge on the field given by the geostatistical model and the data with its associated measurement error.

The Monte-Carlo Markov chains (MCMC) methods seem therefore particularly suited for this problem, as they are known to produce samples of virtually any posterior distribution. Two problems then arise. On one hand, the dimension of the problem is generally the number of grid blocks on which is discretized the random field, often several hundreds of thousands. Therefore, the chain has to be run for an intractable number of iterations to converge on this huge uncertainty space and to achieve an efficient sampling of the posterior. On the other hand, the computational cost of a single fluid flow simulation makes the practitioner wishing to minimize the number of iterations.

MCMC methods have already a long story in history matching, see [22] for a review. More recently, [7] proposed to use the Langevin sampler coupled with Karhunen-Loève (KL) expansion and a multi-scale approach. This algorithm performs a rapid convergence to the stationary regime, but suffers from slow mixing as it performs only a local exploration, as we will illustrate later. In [13], the authors proposed to use adaptive sampling: adjusting the proposal distribution along the chain may improve convergence. However, in this settings, the chain can be trapped into a local mode.

In this work, we first propose a way to reduce the dimension of the inference problem using the KL [19] expansion of geostatistical models. Secondly, we propose a method to improve the global efficiency of the Markov Chain by generating a collection of chains in parallel at different temperatures and allowing them to interact.

The paper is organized as follows. In section 2, we formalize the inverse problem and introduce some notations. In section 3, we show how to reduce the dimension of the problem by the use of a truncated Karhunen-Loève (KL) expansion of the random field of interest and assess the number of components to be kept. In section 4, we first show some negative results given by classical MCMC algorithms on inverse problems through a toy example. We then introduce the principles of interacting Markov Chains. We show on the previous inverse problem example how they provide better mixing while reducing the autocorrelation along the chain, with respect to classical MCMC. In section 5, we show the results of the application of our methodology on a synthetic history matching case. The paper ends with conclusions and perspectives of future work.

2 Notations

In this paper, we consider a Gaussian random field $(X_u)_{u \in \mathbf{U}}$ characterized by its mean μ and covariance Γ , and indexed on \mathbf{U} , the physical domain of the reservoir. The permeability is considered log-normal and it is computed as e^X . The production data is denoted D^* and the measurement error is considered Gaussian with a mean zero and a covariance C_D . The production data are computed for a given random field X by a fluid-flow simulator or forward operator $F(X)$, including the exponential function. With these notations, as given in [28], the posterior distribution is known up to a multiplicative constant independent of X :

$$P(X|D^*) \propto e^{\left(-\frac{1}{2}\|D^* - F(X)\|_{C_D^{-1}}^2 - \frac{1}{2}\|X - \mu\|_{\Gamma^{-1}}^2\right)}, \quad (1)$$

where $e^{\left(-\frac{1}{2}\|X - \mu\|_{\Gamma^{-1}}^2\right)}$ represents the prior probability distribution up to a constant and $e^{\left(-\frac{1}{2}\|D^* - F(X)\|_{C_D^{-1}}^2\right)}$ is termed the likelihood function in the literature of inverse problems. The latter measures the misfit between the production data computed for a given realization X , $F(X)$, and the one observed D^* .

3 Reducing the dimension with Karhunen-Loève expansion

The KL expansion [19] is a parameterization method based on the structural parameters of the random field of interest, namely its covariance operator. It consists in representing the random field in a truncated basis of its covariance operator eigenfunctions. This method has been discussed as long ago as 1976, see [9]. Then, it has been addressed in [23] in a more realistic framework. This method is regaining popularity in reservoir characterization applications, since powerful computing facilities allow to employ it. Recently, such an application has been done in [26], but addressing the problem of choosing the number of components in a

minimalist way and considering the components of a non-Gaussian random field as independent. An interesting work has been done using it in [7] where it has been used conjointly with a MCMC method. We can also cite [6] where it is used in the ensemble Kalman filtering framework. In a previous work [25], we attempted to assess the number of components that has to be used in a fluid-flow application. After a brief summary of the underlying theory, we recall some of our results.

3.1 Brief summary of the underlying theory

The principle of the KL expansion for Gaussian random field is mainly based on the concept of *Reproducing Kernel Hilbert Space* (RKHS). Given X_u , $u \in \mathbf{U}$ a zero mean second order real process with covariance Γ , the RKHS \mathcal{H} of X , or of kernel Γ , is the space of real functions on \mathbf{U} so that :

i) the vectorial space \mathcal{H}_0 generated by the functions $\Gamma(u, \cdot)$ $u \in \mathbf{U}$ is dense in \mathcal{H} ,

ii) for all function h of \mathcal{H} ,

$$h(u) = \langle h(\cdot), \Gamma(u, \cdot) \rangle,$$

where $\langle h(\cdot), g(\cdot) \rangle = \int_{\mathbf{U}} h(u)g(u)du$ is the scalar product on \mathcal{H} .

\mathcal{H} is isomorphic to the space \mathbf{H} generated by the random variables X_u , $u \in \mathbf{U}$. Furthermore, if (ϕ_i) is an orthogonal family of \mathcal{H} , for all family (ξ_i) of independent random variables with law $\mathcal{N}(0, 1)$, the process X defined by:

$$X_u(\omega) = \sum_{i \in I} \xi_i(\omega) \phi_i(u), \quad \omega \in \Omega,$$

is a Gaussian process with covariance Γ .

A judicious choice of the base (ϕ_i) allows to obtain versions with interesting properties. Particularly, the KL's theorem [19] proposes to use the family of $\Gamma(u, \cdot)$ eigenfunctions as an orthonormal family of \mathcal{H} . Indeed, $\Gamma(u, \cdot)$ being positive definite, Mercer's theorem on positive definite kernels ensures existence and uniqueness of this decomposition. The KL theorem says that, if Γ is a continuous covariance on $\mathbf{U} \times \mathbf{U}$, there exists an orthonormal basis (ϕ_i) of \mathcal{H} , constituted by eigenfunctions of $\Gamma(u, \cdot)$. The corresponding basis of \mathbf{H} is the family (ξ_i) of centered uncorrelated random variables. In particular, we have the following representations:

- i) $\xi_i(\omega) = \int_{\mathbf{U}} X_u(\omega) \phi_i(u) du$,
- ii) $X_u(\omega) = \sum_{i \in I} \xi_i(\omega) \phi_i(u)$, $\omega \in \Omega$,
- iii) $\text{var}(\xi_i) = \lambda_i$, where λ_i is the eigenvalue associated with ϕ_i ,
- iv) If X_u is a Gaussian process, the (ξ_i) are independent Gaussian random variables with variance λ_i .

Truncating the sum in ii) above, we can construct approximations:

$$X_u^{(M)} = \sum_{i=1}^M \xi_i(\omega) \phi_i(u) \tag{2}$$

of the true process that minimize the integrated mean squared error on \mathbf{U} : $\varepsilon_M^2 = \int_{\mathbf{U}} E(X_u^{(M)} - X_u)^2 du$, as can be seen in [11]. Indeed, the λ_i being in decreasing order, the first related terms will represent the major part of the global variance. It is worth noticing that the components (ξ_i) are independent if and only if X is Gaussian. In the contrary case, although the (ξ_i) are uncorrelated, they are not independent, and their dependence links need to be studied carefully before to use the KL expansion.

In the following paragraph, we examine the performances of this approximation in terms of variance reproduction.

3.2 Approximation with respect to a statistical criterion

In this paragraph, we show the manner in which a truncated KL expansion of a random field can reproduce the major part of the variance of non-approximated random field. We compute the eigendecomposition of the three following classical geostatistical covariance models, for $(u, v) \in \mathbf{U}$:

1. The exponential model: $\Gamma(u, v) = \sigma^2 e^{-3 \frac{\|u-v\|}{a}}$,

2. The normal model: $\Gamma(u, v) = \sigma^2 e^{-3 \frac{\|u-v\|^2}{a^2}}$,
3. The spherical model: $\Gamma(u, v) = \sigma^2 (1 - 3 \frac{\|u-v\|}{2a} + \frac{\|u-v\|^3}{2a^3}) \mathbb{1}_{\{\|u-v\| < a\}}$, where $\mathbb{1}$ is the indicator function.

As an example, figure 1 presents the rates of variance reproduction in function of the cumulative sum of eigenvalues, for the three models, in two dimensions, with a range $a = H/4$, where H is the size of the field side. We change the discretization step from $H/10$ to $H/60$ in order to assess its effect on the number of components.

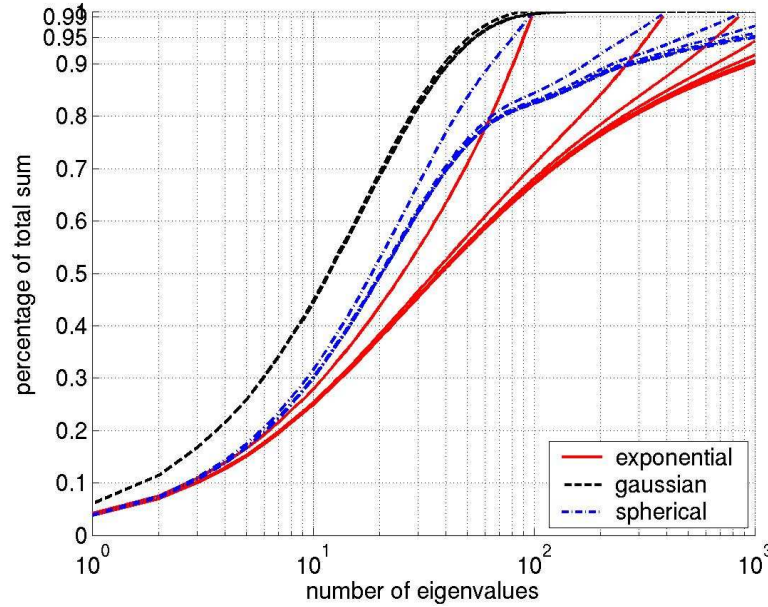


Fig. 1 Cumulative sum of eigenvalues, 2D case, range = $H/4$.

The first thing we can see on this figure is the impact of the discretization step. Indeed, the number of components needed to reproduce, say 90% of the total variance, increases with the number of grid nodes. However, the amplitude of this phenomenon depends on the covariance model, and beyond a certain thinness of the step, this number does not increase anymore.

If we look at the curves corresponding to a discretization step of $H/60$, hence corresponding to a random field discretized on $60 \times 60 = 3600$ grid blocks, we can see that the number of components necessary to reproduce 90% of the global variance for the normal, spherical and exponential models is respectively : 40, 300 and 1000. This number of components is then related to the random field regularity defined by the covariance model.

Moreover, the dimension of the problem of the field also has an impact on the number of components to use. Indeed, increasing the dimension of the random field considered and hence of its covariance operator, will translate itself into an increased number of components to reproduce a given rate of global variance.

We have shown here how a small number of components can represent a large part of the global variance of a random field. However, this criterion is somewhat abstract with respect to our application domain. Therefore, in the next section, we will examine dynamical results obtained by flow simulation on our approximated random fields.

3.3 Conditioning to raw data

Often, we dispose of raw data on the field, *e.g.* at well locations. We explain here briefly how to condition a Gaussian random field X , given its value at two points u and v . In this case, we consider the conditional covariance matrix given by:

$$\Gamma_{M-2,M-2}^c = \Gamma_{M-2,M-2} - \Gamma_{M-2,2} \Gamma_{2,2}^{-1} \Gamma_{M-2,2}^t,$$

where: $\Gamma_{M-2,M-2}$ is the covariance matrix between unconditioned part of the field,

$\Gamma_{M-2,2}$ is the covariance matrix between the two conditioning points and the rest of the field,

$\Gamma_{2,2}$ is the covariance matrix between the two conditioning points, $^{-1}$, t are respectively the inverse and transpose operator.

The conditional mean is given by:

$$\mu^c = \mu + \Gamma_{M-2,2}^t \Gamma_{2,2}^{-1} X_{(u,v)},$$

where $X_{(u,v)}$ is the vector of conditioning data in u and v .

We then have $X|X_{(u,v)} \sim \mathcal{N}(\mu^c, \Gamma^c)$. We then calculate the eigenvalues and eigenvectors of Γ^c .

3.4 Dynamical results

In this paragraph, we consider a 2 dimensional case test with the following properties: the permeability field is modeled by a log-normal random field, with a spherical covariance structure of range 600 m , the mean and variance the logarithm of the field are respectively 3.8 and 1 ; the field size is 2500×2500 m^2 and is discretized on a regular grid of 50×50 blocks ; the thickness of the field is 10 m ; we put two wells on this field: an injector at location (3, 3) and a producer at (48, 48); the permeability field is conditioned with respect to a permeability of 90 md at the well locations ; the porosity is assumed constant at 0.25. The field is assumed to be saturated in oil at time 0. The flow is simulated with *3DSL* [27] during 10000 days with an injection rate at 5000 m^3/day and a pressure of 200 $bars$ at the producer.

We performed a Monte-Carlo study, that is we generated 100 realizations of the permeability field X , conditional to the permeability values at wells location, considering each of them at different truncation orders $X^{(M)}$. We then examined the error made on the water cut (WC) by the approximations at different orders and the true reference field. The water cut is the proportion of water in the oil produced in function of time. Hence, D^* and $F(X^{(M)})$ are both functions of time. The error ε_{WC} is then calculated by the following formula: $\varepsilon_{WC} = \sum_{t=0}^{10000} (F_t(X) - F_t(X^{(M)}))^2$. Figure 2 shows the mean and maximum relative error on the water cut in function of the number of components of the truncation on a logarithmic scale. Furthermore, assuming an allowable error of 2% on the water cut value, we compute a boundary error as the the dot line in red for the mean (a.). We also plot a boundary error of 5% for the maximum as the the dot line (b.).

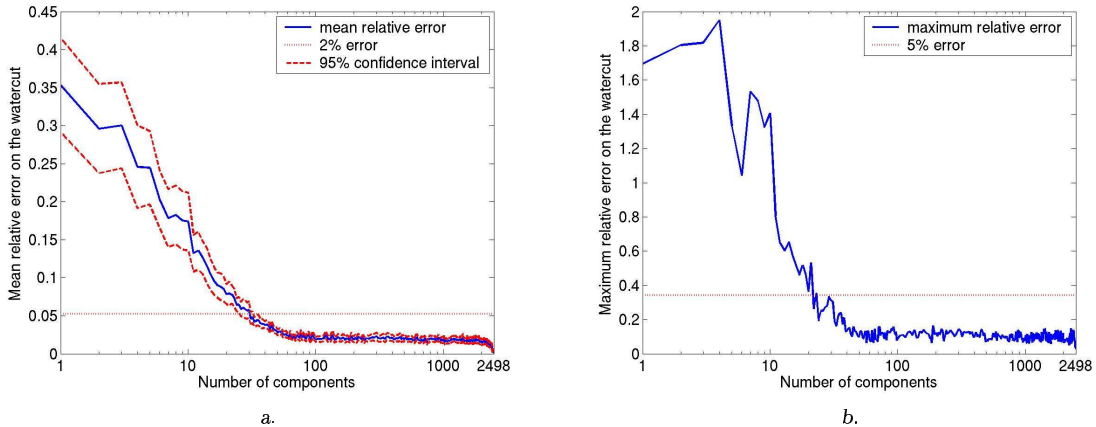


Fig. 2 Mean and maximum relative error on the water cut in function of the number of components.

We can see in figure 2 a. that it suffices, on average, to consider about 30 components to obtain an approximation error on the water cut values below the previously defined confidence level. We also see that it is useless to consider more than 50 components. Indeed, the mean error stabilizes from 50 to 2000 components, then converges quickly toward zero. The maximum relative error in figure 2 b. exhibits the same behaviour. In this example, fulfilling the dynamic criterion requires much less components than that for reproducing 90% of the variance. Indeed, if we turn back at the figure 1, we can see that 30 components correspond to about 60% of the global variance and that 50 components correspond to a bit less than 80% of it. We can then think that only this portion of the variance and the associated eigenfunctions suffices to determine the considered dynamical behaviour of the flow in a permeability field modeled by such a Gaussian random field.

3.5 The eigenvalue problem

The main numerical problem induced by employing the KL decomposition is the computation of the eigenvalues and eigenvectors of the covariance operator. Formally, it takes the form of an homogeneous Fredholm integral equation of the second kind:

$$\int_{\mathbf{U}} \Gamma(u, v) e(u) du = \lambda e(v) \quad (3)$$

For certain classes of stochastic process, there exists an analytical solution to (3), see [11] for some examples. However, for the random fields used in petroleum engineering, (3) needs to be solved numerically.

In a first step, (3) is discretized on the grid of the field. This problem turns then out to be a linear eigenvalue problem, whose dimension is the number of grid nodes used in the discretization. Practically, it can be in the order 10^5 to 10^7 . This is a very challenging problem. Although there exists numerical methods to extract some of the eigenvalues for such high dimensional problems, like [18] (of which there exists now a parallel version), the following idea can be used as a premise:

We consider a gaussian random field X with mean μ and covariance Γ . We cut this field into two blocks X_{B_1} and X_{B_2} . We can then write Γ as follows:

$$\Gamma = \begin{pmatrix} \Gamma_{B_1} & \Gamma_{B_1, B_2} \\ \Gamma_{B_2, B_1} & \Gamma_{B_2} \end{pmatrix}, \quad \mu^t = (\mu_{B_1}, \mu_{B_2})^t$$

And we can simulate a realization of X in the following way:

1. simulate $X_{B_1} \sim \mathcal{N}(\mu_{B_1}, \Gamma_{B_1})$ and get a realization x_{B_1} of X_{B_1} ,
2. then simulate $(X_{B_2} | X_{B_1} = x_{B_1}) \sim \mathcal{N}(\mu_{B_2} + \Gamma_{B_2, B_1}^t \Gamma_{B_1}^{-1} x_{B_1}, \Gamma_{B_2} - \Gamma_{B_2, B_1} \Gamma_{B_1}^{-1} \Gamma_{B_1, B_2})$.

This method can be easily generalized to n blocks and can be then seen as a *sequential block simulation*. There is still a numerical difficulty, as we need to compute the inverse of the matrix Γ_{B_1} , but it is of rather smaller dimension than the whole field. Once computed the conditional block covariance matrices, the KL decomposition can be performed on each one. Therefore, the high dimensional eigendecomposition problem terms to a sequence of low dimensional ones.

We have shown in this section how the dimension of the history matching inverse problem can be drastically reduced by the use of a truncated KL expansion, when the variable of interest is modeled by a second order random field. This method allows to describe approximately a discretized second order random field with a small number of components, while preserving both spatial variability and model uncertainty space. Its assessment on a dynamical flow test shows that only a few components governs the fluid flow behaviour. This method is then of great interest for solving the history matching inversion problem. Its assessment on a dynamical flow test shows that only a few of those components governs the flow behaviour.

However, different parameters can influence the number of components such as the range, the regularity of the covariance, and the dimension of the random field. Hence, a preliminary study has to be performed each time we want to use the decomposition, in order to choose the number of components with respect to a given criterion. Otherwise, this number can be determined according to the available computational resources.

Note also that a fundamental property of the KL expansion for Gaussian random field is the independence of its random components. Once the desired eigenvalues and associated eigenfunctions are known, it is straightforward to generate an approximate random field, that is of capital importance for applying MCMC algorithm.

4 MCMC for inverse problem

In this section, we first recall the main principles of MCMC. Secondly, we illustrate the drawbacks of using MCMC for inverse problems through a toy example. We finally propose a strategy to surpass these drawbacks, introducing the principles of parallel interacting Markov Chains.

4.1 General principle

MCMC methods, introduced by Metropolis et al.[20], is a popular method for generating samples from virtually any distribution π defined on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. In particular, there is no need for the normalising constant to be known and the space $\mathcal{X} \subseteq \mathbb{R}^d$ (for some integer d) on which it is defined can be high dimensional. The method consists in simulating an ergodic Markov chain $\{X_n, n \geq 0\}$ on \mathcal{X} with transition probability P such that π is a *stationary* density for this chain, *i.e.* $\forall A \in \mathcal{B}(\mathcal{X})$:

$$\int_{\mathcal{X}} P(x, A) \pi(x) dx = \pi(A) \quad (4)$$

Such samples can be used *e.g.* to compute integrals

$$\pi(g) = \int_{\mathcal{X}} g(x) \pi(x) dx, \quad (5)$$

estimating this quantity by

$$S_n(g) = \frac{1}{n} \sum_{i=1}^n g(X_i), \quad (6)$$

for some $g : \mathcal{X} \rightarrow \mathbb{R}$. We illustrate the principles of MCMC with the Metropolis-Hastings (MH) update. It requires the choice of a *proposal distribution* q . The role of q consists in proposing potential transitions for the Markov chain. Given that the chain is currently at x , a candidate y is accepted with probability $\alpha(x, y)$ defined as:

$$\alpha(x, y) = \begin{cases} \min \left(1, \frac{\pi(y) q(x, y)}{\pi(x) q(y, x)} \right) & \text{if } \pi(x) q(x, y) > 0 \\ 1 & \text{else.} \end{cases} \quad (7)$$

Otherwise, it is rejected and the Markov chain stays at its current location x . The transition kernel P of this Markov chain takes the form, for $(x, A) \in \mathcal{X} \times \mathcal{B}(\mathcal{X})$:

$$P(x, A) = \int_A \alpha(x, y) q(x, y) dy + \mathbb{1}_A(x) \int_{\mathcal{X}} (1 - \alpha(x, y)) q(x, y) dy, \quad (8)$$

The Markov chain defined by P is reversible with respect to π and therefore admits π as invariant distribution. Conditions on the proposal distribution q that guarantee irreducibility and positive recurrence are easy to meet and many satisfactory choices are possible. However, we will see in the next paragraph that the practice shows difficulty, particularly for non-linear inverse problems purposes, where the posterior can be highly multi-modal.

4.2 Performances of classical MCMC samplers on a toy example

Three important pitfalls exist in the practice of MCMC. First, MH algorithms such as the symmetric increments random-walk sampler, the sequential independent sampler that is changing at each step a unique component of the field, or the Langevin sampler, suffer from slow mixing or even lack of convergence; especially for a small number of iteration. The second drawback of classical MCMC is the weak rate of acceptance, when we compute the empirical estimate of (7) and it generally worsens when the dimension increases. Particularly in history matching, where one iteration has a high computational cost, one wishes to have a good acceptance rate when using MCMC. The third drawback is the autocorrelations length along the chain: as the realizations are drawn from a Markov Chain, they are dependent one from the other. The variance of the quantities (6) we want to estimate are thus bigger than the empirical variance of the sample. In fact, the longer are the autocorrelations, the bigger will be the variance.

We now describe some classical MCMC samplers and set evidence on the drawbacks of using MCMC

methods in inverse problems on a toy example. Our toy inverse problem is characterized by the following function:

$$\begin{aligned} F : \mathbb{R}^2 &\mapsto \mathbb{R} \\ X = (X_1, X_2) &\rightarrow 2X_1^2 + X_2^2, \end{aligned} \quad (9)$$

given the following prior on X :

$$X \sim \mathcal{N}\left(0, \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}\right) \quad (10)$$

Given a particular realization of X , written $X^* = (1.514, 1.335)$; we assume that we observe $D^* = F(X^*)$ with an error $\varepsilon \sim \mathcal{N}(0, 0.5)$. We plot the posterior density in figure 3 a. and b. from two different viewpoints. We can see its support envelopes the ellipse $\{X \in \mathbb{R}^2 | F(X) = D\}$ and it presents two distinct modes.

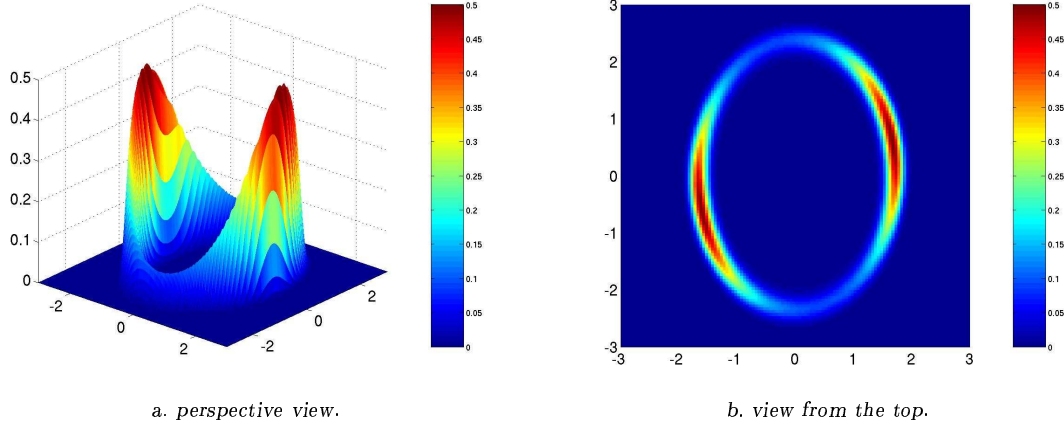


Fig. 3 Posterior distribution.

To solve this problem, we try four different MH samplers:

1. the independent sampler (IMH): $q(x, y) = q(y)$, where q is given by (10),
2. the symmetric increments random-walk sampler (SIMH): $q(x, y) = q(|y - x|)$, where q is the multivariate distribution $\mathcal{N}(0, hI)$, where I is the identity matrix and $h = 0.1$,
3. the Langevin sampler (LMH): assume that π is differentiable on \mathcal{X} , q takes the form:

$$q(x, y) \sim \mathcal{N}\left(x + \frac{h^2}{2} \nabla \log(\pi(x)), h^2 I_d\right), \quad (11)$$

where h is chosen 0.1.

4. The adaptive algorithm of [12] (ASIMH): In this algorithm, y is proposed according to $q_{\theta_n}(x, \cdot) = \mathcal{N}(x, \Gamma_n)$, where $\theta = (\mu, \Gamma)$. We also consider a non-decreasing sequence of positive stepsizes $\{\gamma_n\}$, such that $\sum_{n=1}^{\infty} \gamma_n = \infty$ and $\sum_{n=1}^{\infty} \gamma_n^{1+\delta} < \infty$ for some $\delta > 0$. In practice we use: $\gamma_n = 1/n$, as suggested in [12]. The parameter estimation algorithm take the following form:

$$\begin{aligned} \mu_{n+1} &= \mu_n + \gamma_{n+1} (X_{n+1} - \mu_n), \quad n \geq 0, \\ \Gamma_{n+1} &= \Gamma_n + \gamma_{n+1} \left((X_{n+1} - \mu_n)(X_{n+1} - \mu_n)^T - \Gamma_n \right), \end{aligned} \quad (12)$$

We can notice here that the chain generated by the adaptive algorithm is no longer homogeneous, but it can be proved (see [12] and [4] and [2] in a more general framework) that it has the correct ergodic properties. The idea of adaptive sampling is to improve the proposal efficiency, making it as close as possible to the posterior density. Practically, the parameters of the proposal are not updated at each iteration. We ran this algorithm with a first update of the parameters after 50 iterations and then an update every 10 iterations.

We ran all four algorithms for 10000 iterations, which represent already an important number of iterations for history matching applications. We compare the results given by those four algorithms for the inverse problem defined above. To do this, we introduce three different criteria:

1. the kernel density estimate (KDE) of the posterior in figures 5, 6, 7, 8, to be compared to figure 4. The latter represents the KDE of the posterior constructed from an i.i.d. sample of size 10000 of the true posterior. In the whole paper, we consider the bandwidth of the KDE to be chosen according to the sample size,
2. the empirical acceptance rate in figure 9,
3. the mean autocorrelations along the chain in figure 10.

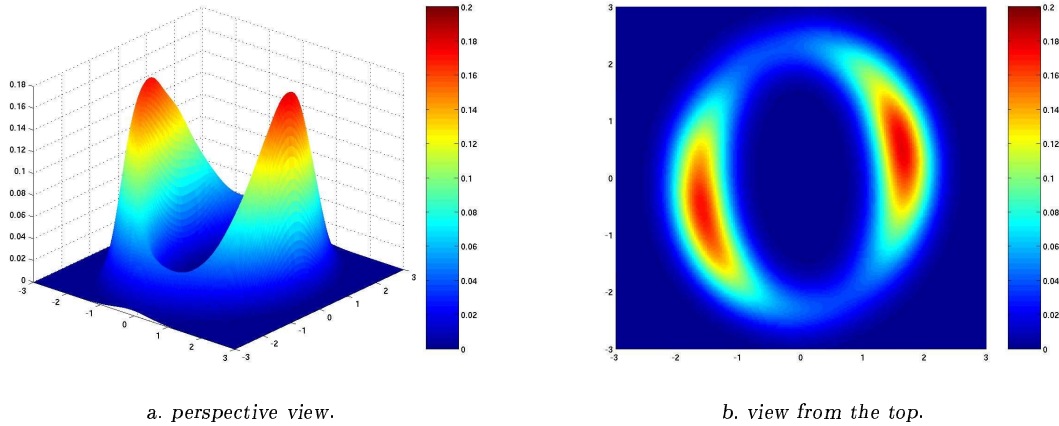


Fig. 4 Kernel density estimate of the posterior constructed on an i.i.d. sample of size 10000.

In figures 5, 6, 7 and 8, we can see the KDE computed from the states generated by all four chains. It is of great interest to remark that the SIMH and the LMH only perform a local sampling of the posterior (figures 6, 7), missing one of the two modes of the posterior. Indeed, their KDE do not exhibit the two modes of the posterior. Only the spannings performed by the IMH and the ASIMH are correct as they charge the entire ellipse. Particularly, the KDE constructed from the IMH sample is correct (figure 5). The one corresponding to the ASIMH sample, in figure 8 is really poor. It is probably due to the fact that we considered the entire sample. Indeed, it is possible that as the first updates of the parameters occur, the algorithm has only spanned a small portion of the space and then it makes appear artefacts due to the use of *bad* parameters. Indeed, the parameters estimates are really slow to converge and it conducts to poor sampling in the beginning of the algorithm. This behaviour makes this algorithm improper to our application domain.

Furthermore, the ASIMH and the IMH exhibit the poorest final empirical acceptance rate (figure 9), around 0.1, meaning that only 10% of the moves are accepted for this relatively simple problem. Conversely, the SIMH and IMH show good performances regarding this criterion, especially the LMH with a final empirical acceptance rate greater than 0.95. However, those two chains show poor performances with respect to autocorrelation length (defined as the first time beyond which the autocorrelation is less than 0.05, the dot line in figure 10), as they exhibit the longest autocorrelation length, unlike IMH and ASIMH.

By looking at these figures, the practitioner has to choose between good mixing properties with short autocorrelation but really poor acceptance rates and bad mixing with long autocorrelation and good acceptance rates. Furthermore, the effect of an increasing dimension is not addressed here but it is known to worsen the properties of the chain. Especially, the IMH algorithm is known to fail in high dimension, although it performed relatively well here. In order to improve the mixing properties of the chains without increasing the computational cost of the algorithm, we now introduce the principle of parallel interacting Markov Chains (IMC).

4.3 Parallel Interacting Markov Chains

The principle of making interact Markov Chains first appears in [10] under the name parallel tempering (PT). It has been mostly applied in physicochemical simulations, see [8] and references therein. It is known in the literature under different names such as: exchange monte-carlo, metropolis coupled-chain, see [15] for a review. The principle of PT is to simulate a number $(K + 1)$ of replica of the system of interest by MCMC, each at a different temperature, in the sense of the simulated annealing, and to allow the chains to exchange information, swapping their current state. The high temperature systems are generally able to sample large

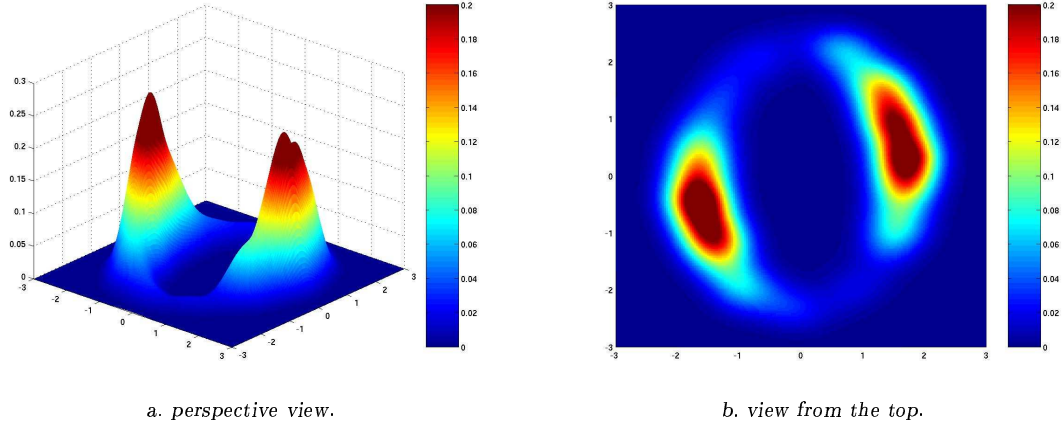


Fig. 5 Kernel density estimate of the posterior constructed on the states generated by the IMH algorithm.

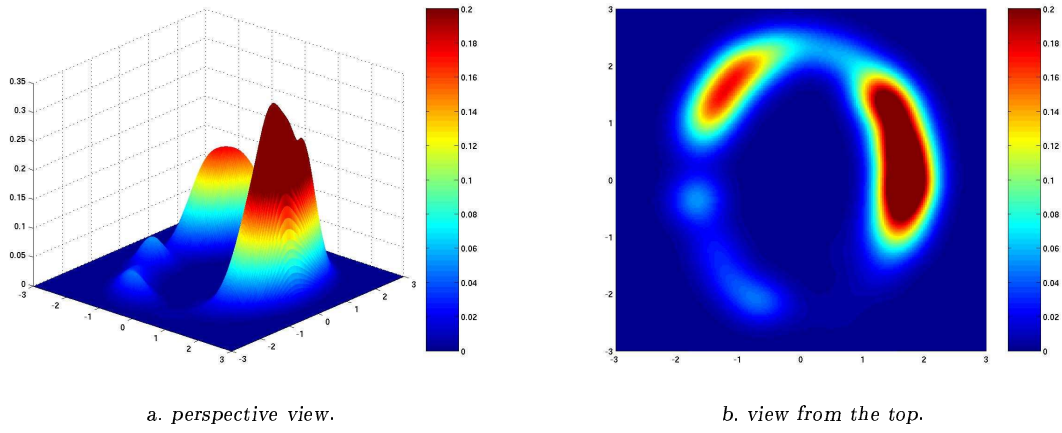


Fig. 6 Kernel density estimate of the posterior constructed on the states generated by the SIMH algorithm.

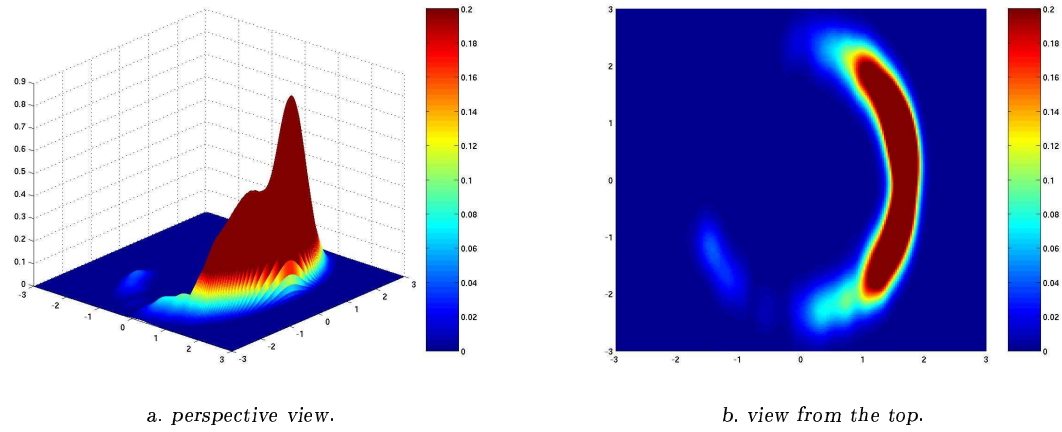


Fig. 7 Kernel density estimate of the posterior constructed on the states generated by the LMH algorithm.

volumes of state space, whereas low temperature systems, whilst having precise sampling in a local region of state space, may become trapped in local energy minima during the timescale of a typical computer simulation. Parallel tempering achieves good sampling by allowing the systems at different temperatures to exchange their state. Thus, the inclusion of higher temperature systems ensures that the lower temperature systems can access a representative set of low-temperature regions of state space.

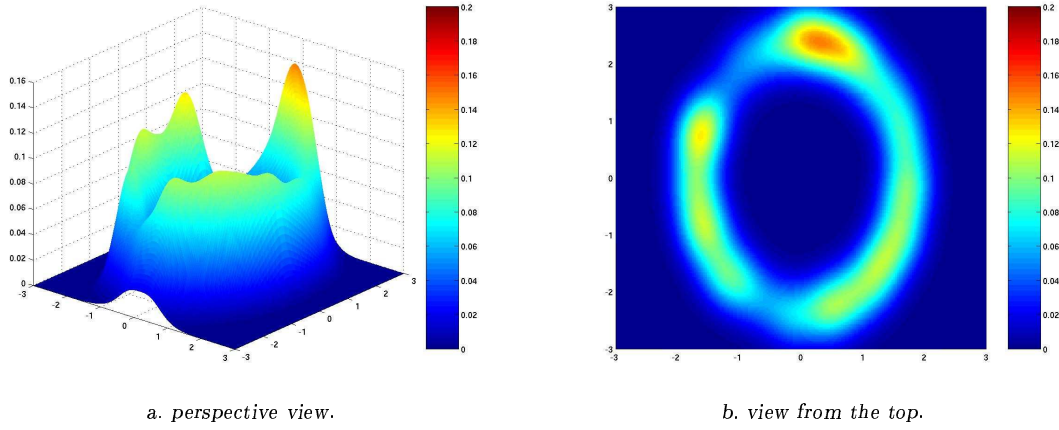


Fig. 8 Kernel density estimate of the posterior constructed on the states generated by the ASIMH algorithm.

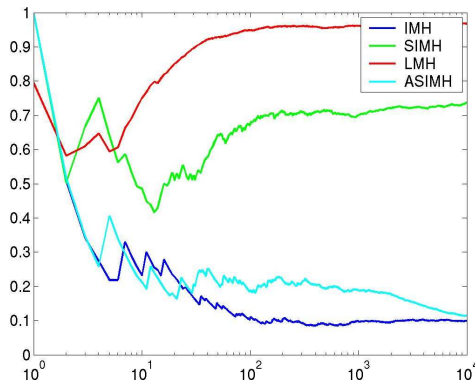


Fig. 9 Empirical acceptance rates along the chain for all four algorithms, logarithmic scale.

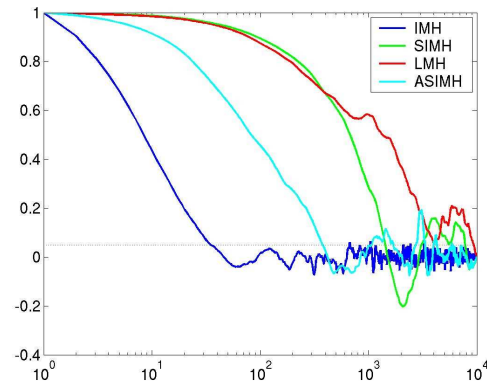


Fig. 10 Autocorrelations along the chain for all four algorithms, logarithmic scale.

Simulation of $(K + 1)$ replicas, rather than one, requires on the order of $(K + 1)$ times more computational effort. This *extra expense* of PT is one of the reasons for the initially slow adoption of the method. Eventually, it became clear that a PT simulation is more than $(K + 1)$ times more efficient than a standard, single-temperature Monte Carlo simulation. This increased efficiency derives from allowing the lower temperature systems to sample regions of state space that they would not have been able to access, even if regular sampling had been conducted for a single-temperature simulation that was $(K + 1)$ times as long. It is also worth noticing that PT can make efficient use of large CPU clusters, where different replicas can be run in parallel, unlike classical MCMC sampling that are sequential methods. An additional benefit of the PT method is the generation of results for a range of temperatures, which may also be of interest to the investigator. It is now widely appreciated that PT is a useful and powerful computational method.

More recently, some researchers in the statistical community took attention on PT and more generally on interacting Markov Chains. They propose a general theoretical framework and new algorithms in order to improve the exchange information step addressed above. Two main algorithms drawn our attention: the equi-energy sampler (EES) of [17] and the population importance-resampling MCMC sampler (PIR) of [1] so called in [3], which allows to go back in the history of the chain. More precisely, these two last algorithms are based on self interacting approximations of non-linear Markov kernels, defined by Andrieu et al. [1]. We now describe these methods in our context.

4.3.1 Description of the algorithms

We first define the family $\{\pi^{(l)}, l = 0..K\}$ of distributions we want to simulate from, such that:

$$\pi^{(l)}(x) \propto e^{-E_l(x)}, \quad (13)$$

where $E_l(x) = \frac{E(x)}{T_l}$ and $E(x) = \frac{1}{2}\|D^* - F(x)\|_{C_D^{-1}}^2 + \frac{1}{2}\|x - \mu\|_{\Gamma^{-1}}^2$ is the energy of the system and T_l is the temperature at which it is considered. The T_l satisfy: $T_0 = 1 < T_1 < \dots < T_K < +\infty$, so that $\pi^{(0)} = P(X|D^*)$. These distributions are so a family of *tempered* versions of $P(X|D^*)$. We will also talk of *tempered* energies to denote the E_l . The parallel algorithms aim to simulate from:

$$\Pi(x) = \prod_{l=0}^K \pi^{(l)}(x), \quad (14)$$

allowing exchanges between states at different temperatures. Dilatate versions of $\pi^{(0)}: \pi^{(1)}, \dots, \pi^{(K)}$ are easier to simulate. Thus they can provide information on $\pi^{(0)}$.

Different strategies can be adopted to exchange information between chains at adjacent temperatures. We denote by $x = (x^{(0)}, \dots, x^{(K)}) \in \mathcal{X}^{K+1}$ the current state of the chain. For $l = 0, \dots, K-1$, we define the importance function:

$$r^{(l)}(x) = e^{-(E_l(x) - E_{l+1}(x))} \quad (15)$$

The method can be formalized by defining the following kernel P_n at time n , given all the previous states $x_{0:n-1} = (x_0, \dots, x_{n-1})$ and for $A_0 \times \dots \times A_K \in \mathcal{B}(\mathcal{X}^{K+1})$:

$$P_n(x_{0:n-1}; A_0 \times \dots \times A_K) = P^{(K)}(x^{(K)}, A_K) \prod_{l=0}^{K-1} P_{x_{0:n-1}}^{(l)}(x^{(l)}, A_l), \quad (16)$$

where for $x_{0:n-1}^{(l+1)} \in \mathcal{X}^n$, $x^{(l)} \in \mathcal{X}$ and $A \in \mathcal{B}(\mathcal{X})$:

$$P_{x_{0:n-1}}^{(l)}(x^{(l)}; A) = \theta P^{(l)}(x^{(l)}, A) + (1 - \theta) \int_{\mathcal{X}} \nu_{x_{0:n-1}}^{(l)}(x^{(l)}, dy) T^{(l)}(y, x^{(l)}; A), \quad (17)$$

and

$$\nu_{x_{0:n-1}}^{(l)}(x^{(l)}, dy) = \frac{\sum_{i=0}^{n-1} \omega_{n,i}^{(l)}(x^{(l)}, x_i^{(l+1)}) \delta_{x_i^{(l+1)}}(dy)}{\sum_{i=0}^{n-1} \omega_{n,i}^{(l)}(x^{(l)}, x_i^{(l+1)})} \quad (18)$$

and in the algorithms considered here $T^{(l)}$ will take the following form:

$$T^{(l)}(y, x^{(l)}; A) = \min\left(1, \frac{r^{(l)}(y)}{r^{(l)}(x^{(l)})}\right) \mathbb{1}_A(y) + \left(1 - \min\left(1, \frac{r^{(l)}(y)}{r^{(l)}(x^{(l)})}\right)\right) \mathbb{1}_A(x^{(l)}) \quad (19)$$

In other words, at time step n , at temperature T_l , with probability θ , a classical MH move will be performed. Else, with probability $(1-\theta)$, an exchange move will be proposed. It consists in choosing a state y among $x_{0:n-1}^{(l+1)}$ with the weights $\omega_{n,i}^{(l)}(x^{(l)}, x_i^{(l+1)})$. This move is then accepted with probability $\min\left(1, \frac{r^{(l)}(y)}{r^{(l)}(x^{(l)})}\right)$. The new state is chosen according to $\nu_{x_{0:n-1}}^{(l)}$, that can be viewed as an importance sampling estimate of $\pi^{(l)}$ with the instrumental law $\pi^{(l+1)}$.

The three algorithms (PT, EES, PIR), considered in this article can be written in this framework, and differ only from the formulation of the weights $\omega_{n,i}^{(l)}$, for some $(y, z) \in \mathcal{X}^2$:

- for the PT algorithm, we have :

$$\omega_{n,i}^{(l)}(y, z) = \mathbb{1}_{i=n-1},$$

it is only possible to go to the current state of the chain at the adjacent higher temperature,

- for the EES algorithm, given a sequence of energy levels $E_0 < E_1 < \dots < E_K < E_{K+1} = \infty$ defining a partition: $\mathcal{X} = \bigcup_{l=0}^K \mathcal{X}_l$ of energy rings: $\mathcal{X}_l = \{x \in \mathcal{X} : E_l < E(x) < E_{l+1}\}$, and the function $I(x) = l$ if $x \in \mathcal{X}_l$ the $\omega_{n,i}$ take the form:

$$\omega_{n,i}^{(l)}(y, z) = \mathbb{1}_{\mathcal{X}_{I(y)}}(z),$$

that is, the new state will be taken uniformly among the states $x_{0:n-1}^{(l+1)}$ of the chain at temperature T_{l+1} that are in the same energy ring as the current state,

– for the PIR algorithm, the weights $\omega_{n,i}$ take the following form:

$$\omega_{n,i}^{(l)}(y, z) = r^{(l)}(z),$$

that is we obtain the new state by resampling from $x_{0:n-1}^{(l+1)}$ with the weights ω .

The main idea behind the two last algorithm is that the kernel defined in (17) will converge towards the following limiting kernel:

$$P_{x_{0:n-1}}^{(l+1)}(x^{(l)}; A) = \theta P^{(l)}(x^{(l)}, A) + (1 - \theta) R^{(l)}(x^{(l)}, A), \quad (20)$$

where $R^{(l)}$ is a MH kernel, whose proposal distribution is given by:

- $Q_{PIR}^{(l)}(x, dy) = \pi^{(l)}(dy)$ for the PIR algorithm,
- $Q_{EES}^{(l)}(x, dy) \propto \pi^{(l+1)}(y) \mathbb{1}_{\mathcal{X}_I(x)}(y) \lambda(dy)$ for the EES.

Obviously the convergence towards $R^{(l)}$ will not be achieved in the time of the simulation, but its approximation at time n will help to sample from the posterior, particularly to span a larger part of the state space. Finally, it is worth noticing that for all 3 algorithms, we can use the samples of all the chains, reweighting them by the following importance weights:

$$\eta^{(l)}(x) = e^{-(E_0(x) - E_l(x))}, \quad (21)$$

in order to compute estimates of $I_g = E_{\pi_0}[g(X)]$, for some g . Hence, the estimate \hat{I}_g , after N iteration of the algorithm will take the form:

$$\hat{I}_g = \sum_{l=0}^K \frac{\sum_{i=0}^N \eta^{(l)}(x_i^{(l)}) g(x_i^{(l)})}{\sum_{i=0}^N \eta^{(l)}(x_i^{(l)})}. \quad (22)$$

It has been shown numerically in [17] that using the reweighted entire sample will provide better estimates than using only $x_{0:N}^{(0)}$.

Concerning the choice of the parameters, some heuristic rules exist and are discussed in e.g. [15] for the PT algorithm and in [17] for the EES. However, the choice depends mainly on the problem addressed and there is no general recipe to tune the parameters. We will explain how we choose them for the history matching application in paragraph 5.2.

We now describe some numerical results for these 3 algorithms.

4.3.2 Numerical results

We compare the results given by these three algorithms on the problem defined in paragraph 4.2. We consider 4 different temperatures ($T_3 = 50$, $T_2 = 13.5722$, $T_1 = 3.6841$, $T_0 = 1$) and we set the probability of exchange θ at 0.2. At each temperature, we use a SIMH sampler with a variance varying according to the temperature, namely 0.1 at T_0 and $0.1\sqrt{T_l}$ at T_l , $0 < l \leq 3$.

We now compare the results given by those three algorithms for the inverse problem defined in 4.2. We do not reproduce here each of three different criteria we used previously. Indeed, in terms of empirical acceptance rate, for each chain, our three algorithms behave like the SIMH. We recall the two criteria, defined in 4.2, that we use there :

1. the mean autocorrelations for each parallel chain along the chain in figures 11, 12 and 13,
2. the KDE of the posterior in figures 14, 15, and 16 to be compared to figure 4 . Again, we used a bandwidth chosen according to the sample size. We constructed these estimates considering the whole sample, that is, including the states of all four chains, weighting them by the weights (21).

When we look at figures 11, 12 and 13, we can see that these methods provide the major advantage of reducing autocorrelations along the chain with respect to single MCMC. For each of 3 algorithms, we represent the autocorrelations along each of 4 chains. We can see that the autocorrelations along the chain at T_0 are systematically reduced with respect to the classical SIMH algorithm (see figure 10), although the proposal variance has been chosen equally to be 0.1. This phenomenon is due to the exchange steps that allow

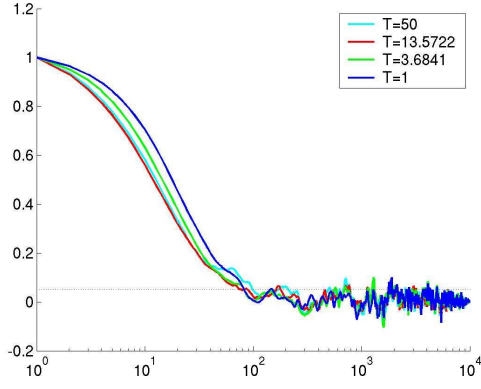


Fig. 11 Autocorrelations along the chain for the PT algorithm, logarithmic scale.

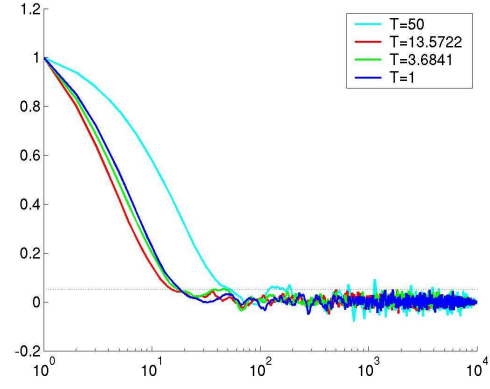


Fig. 12 Autocorrelations along the chain for the EES, logarithmic scale.

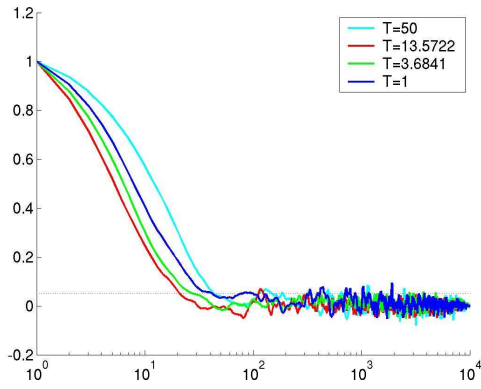


Fig. 13 Autocorrelations along the chain for the PIR algorithm, logarithmic scale.

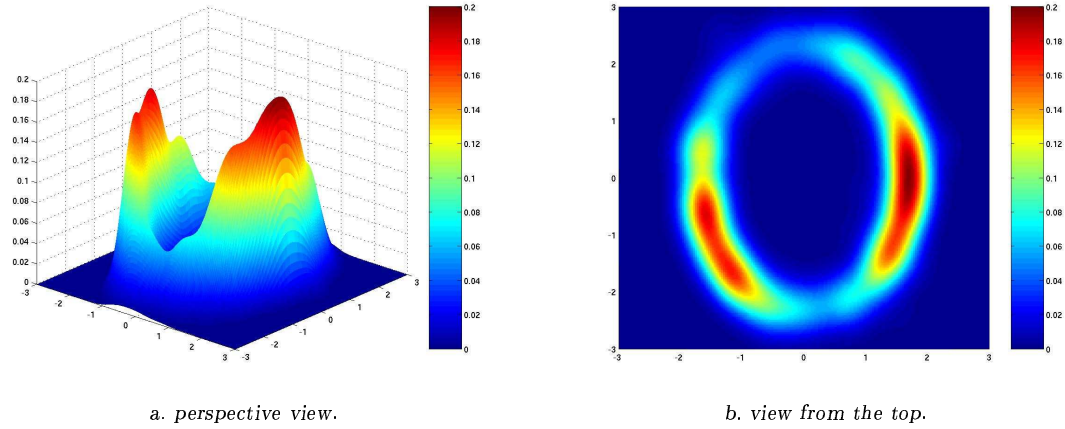


Fig. 14 Kernel density estimate of the posterior constructed on the states generated by the PT algorithm.

regenerations of the chains. In other words, each piece of chain between two exchange steps is independent from one another, which induces the reduction of the autocorrelations.

In figures 14, 15 and 16, we represent the resulting KDE for all three algorithm. We can see the ability of these methods to sample efficiently multi-modal distribution: the density estimates clearly exhibit the two modes of the posterior distribution.

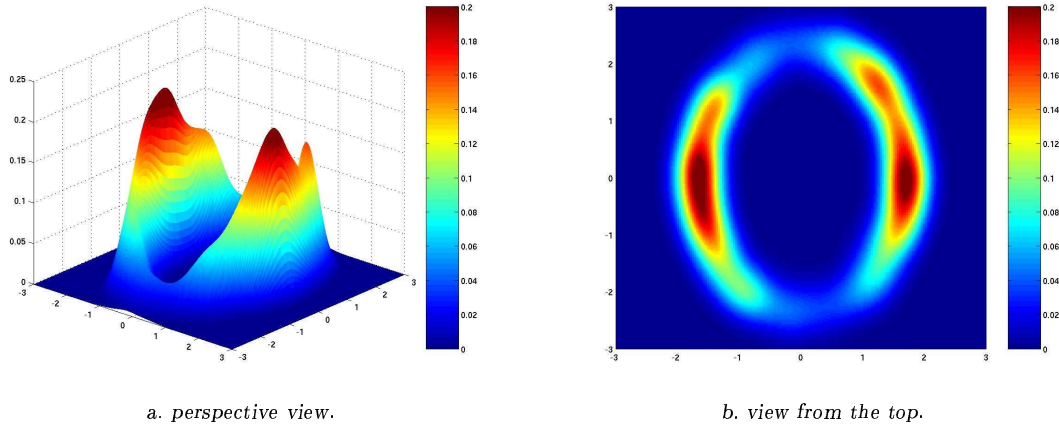


Fig. 15 Kernel density estimate of the posterior constructed on the states generated by the EES.

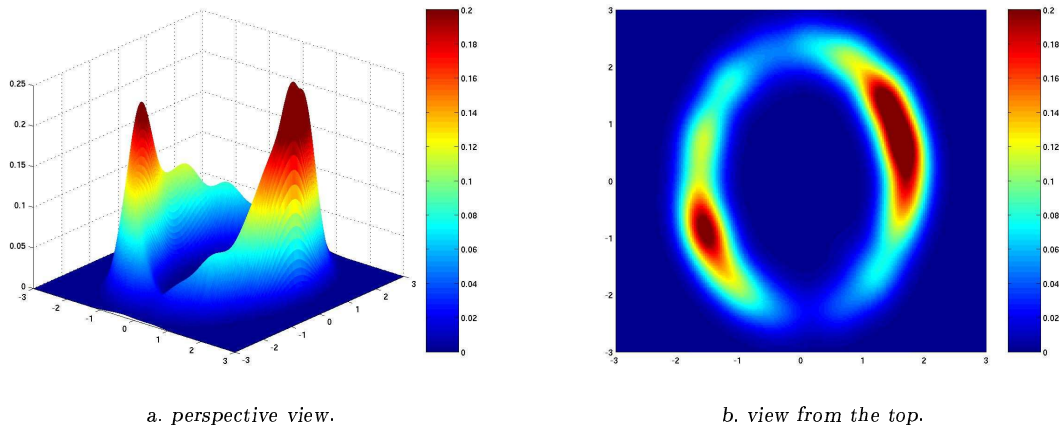


Fig. 16 Kernel density estimate of the posterior constructed on the states generated by the PIR algorithm.

Finally, we have shown on this toy example the advantage of using interacting Markov Chains: they improve the browsing of the support of the distribution while reducing the autocorrelations, comparatively to classical single MCMC sampling. These performances are realized without increasing the computational cost, as these algorithms are easily parallelizable. In fact, at fixed iteration number, the parallel algorithms are even faster, as an exchange step counts for one iteration, without computing the forward operator. These methods are then particularly suited for solving inverse problems where no analytical expression for the forward operator exists, like the history matching problem. More precisely, the PIR and the EES algorithm are more powerful than the PT, as they allow exchanges with the past of the chains, accounting for all the information learned until the iteration n . In the next section, we apply the PIR algorithm to a reservoir characterization synthetic problem.

5 Application to history matching

In this section, we describe the results given by the PIR algorithm on a synthetic reservoir characterization problem. Indeed, we think it is the most suited algorithm for our problem. Considering that the lower is the temperature, the slower the chain will enter the stationary regime, we can remark that the PIR does not need the chains to be in stationary regime before to allow exchanges, contrarily to the EES algorithm. Indeed, in the EES, the exchange proposal is made in the same energy ring as the current state. Then, if the chain of interest has not reached the stationary regime and is still at high energy levels, the exchange proposal will be in the same energy ring as the current state. Therefore, it will not help to attain stationary regime. Conversely, the PIR proposes exchange proposals according to an importance sampling step, constructed on

the states generated at the higher adjacent temperature. The proposals are then more likely at low energy levels. Therefore, we choose to apply here the PIR algorithm, described in paragraph 4.3.1.

5.1 Description of the case

The prior geostatistical model of the field for this problem and the parameters of the fluid-flow simulator are the same as in 3.4. Given a reference realization of the field X , computed with all its 2498 components, and its water cut calculated on 3000 days, we attempt to condition the geostatistical model X to the water cut D^* . In other words, we try to produce an i.i.d sample of (1) as explained in section 2. However, we use a truncated KL expansion (2) with $M = 50$ components to represent the field. Hence, the dimension of the inference problem is reduced. The formulation of the posterior is then the following:

$$P(X^{(M)}|D^*) \propto e^{\left(-\frac{1}{2}\|D^* - F(X^{(M)})\|^2 - \frac{1}{2}\|X^{(M)} - \mu\|_{\Gamma(M)^{-1}}^2\right)}, \quad (23)$$

where $\Gamma_{(M)} = \Phi_{(M)} \Lambda \Phi_{(M)}^t$,
 $\Phi_{(M)}$ is the matrix $L \times M$ whose column vectors are the $\phi_i(x)$,
 Λ is the diagonal $M \times M$ whose diagonal components are the λ_i .

Here, we consider the production data water cut. We recall here that the water cut is the proportion of water in the oil produced, in function of time. Hence, D^* and $F(X^{(M)})$ are both functions of time. The covariance of the measurement error on the water cut is assumed to be I , where I is the identity matrix. The exponent of the likelihood term in (23) is written:

$$\|D^* - F(X^{(M)})\|^2 = \sum_{t=0}^{t=3000} \left(D_t^* - F_t(X^{(M)})\right)^2$$

5.2 Methodology

In order to sample from (23), we implement the PIR algorithm. In this section, inspired by practical considerations found in [15] and [17] and our own experience, we explain how we choose the different parameters of the algorithm. Particularly, we will focus on the following four points :

1. the highest temperature to choose,
2. the kernel to choose as a function of the temperature,
3. the number of chains,
4. the probability of proposing information exchange between chains.

We first launch 200 preliminary runs in order to choose the maximum temperature we will use. We represent the histogram of energies in figure 17. In figure 18, we also show the corresponding curves of water cut together with the black thick curve representing the reference water cut. Moreover, we represent in this figure the 0 to 3000 days period to be matched as well as the 3000 to 5000 days period. This last part of the curve is computed to assess the prediction ability of the sample. We can see on this figure that all the curves are far away from the reference water cut and are really poor for prediction purposes.

We can see in figure 17 that the energy are mainly distributed between 0 and 5000 with an average of about 3000 (the minimum energy of the sample being about 800). As proposed in [15], we choose first a maximal temperature. We consider the energy term in (23):

$$E(X) = \frac{1}{2}\|D^* - F(X^{(M)})\|^2 + \frac{1}{2}\|X^{(M)} - \mu\|_{\Gamma(M)^{-1}}^2. \quad (24)$$

The mean of the second term of (24) is:

$$\mathbb{E} \left(\frac{1}{2}\|X^{(M)} - \mu\|_{\Gamma(M)^{-1}}^2 \right) = \mathbb{E} \left(\frac{1}{2} \sum_{i=1}^M \xi_i^2 \right) = \frac{M}{2},$$

when the ξ_i , defined in paragraph 3.1, are considered i.i.d centered reduced Gaussian. Formally, the posterior distribution of the ξ_i is not Gaussian, as F is not linear. Nevertheless, we only try here to evaluate an order

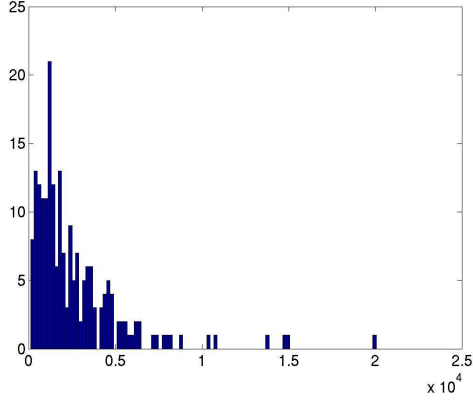


Fig. 17 Histogram of energies of 200 random realizations.

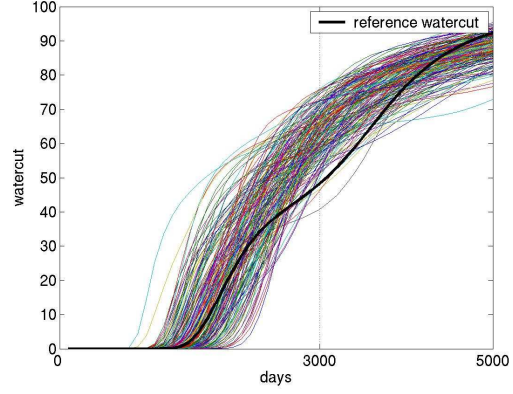


Fig. 18 Water Cut curves of the sample.

of magnitude of the mean energy of the posterior, in order to construct the temperature ladder. In the same time, the first term of (24) has to be near zero for a realization of (23). We deduce that the mean of (24) for the conditioned realizations will be about $M/2$, that is 25 here. As the average energy of the 200 Monte-Carlo sample has been found 5000, we choose the highest temperature to be 100, to ensure a sufficient transition acceptance rate for the chain at the highest temperature.

As already claimed, the idea of these methods is to improve the mixing of the chain. Then, according to the temperature, we have to choose kernels that will make effective this assumption. At the highest temperature, larger moves tends to be accepted, even though the energy level reached is not as low as the one finally aimed. Thus, it is of great interest to use a fast mixing kernel that cannot be used at lower energy levels. For this application, a first run using a SIMH with very large moves at highest temperature, as proposed in [17], led to very poor results, generating only extreme values with very high energy levels at the highest temperatures. Thus, the exchange propositions were systematically rejected. We then chose an independent sampler at this temperature.

Conversely, at low temperatures, it is of interest to employ a kernel with good local properties like the Langevin sampler or a SIMH with small steps. This kind of kernel, will perform a good local exploration of the posterior. The point is then to choose the intermediate kernels, between the highest and the lowest temperature levels. Increasing the step of a SIMH sampler according to the temperature, as proposed in [17] seemed to be a good idea. However as said previously, it failed. The difficulty is then to choose kernels that progressively worsen their mixing properties, while increasing local properties, when descending the temperature ladder. In a high dimensional setting like ours, one can change the number of components that are affected at each iteration. For example, we can choose a kernel that modifies all components at each iteration at the highest temperature, and a kernel that modify only one at the lowest.

Furthermore, these considerations about kernels are closely related to the number of chains to use. Particularly, it is important not to employ too numerous chains. Indeed, using too much chains will slow the input of information from the highest temperature levels to the lowest. Conversely, the number of chains has to be large enough to allow them to exchange information at a good rate: the histograms of the tempered energies of two chains at adjacent temperatures need to possess a sufficient overlapping interval to allow exchanges.

Here, after some numerical experiments, we chose to use 5 different temperatures. We construct the temperature ladder distributing geometrically the temperatures between $T_0 = 1$ and $T_4 = 100$. The choice of a geometric distribution of the temperatures is a classical one in the parallel tempering literature, see e.g. [8]. If the number of chains is sufficient, it allows generally a good overlapping of the histograms of the tempered energies. Namely, we take $T_l = T_0 \left(\frac{T_4}{T_0} \right)^{l/4}$ for $l = 1, 2, 3$. Hence, we use the the following temperature ladder:

$$T_0 = 1.000 < T_1 = 3.162 < T_2 = 10.000 < T_3 = 31.623 < T_4 = 100.000$$

Thus, we simulate 5 Markov chains $(X^{(l)})$ at the temperature $T^{(l)}$. At T_0, T_1, T_2 , we simulate from a SIMH algorithm with a step variance $0.1\sqrt{T_l}$, affecting respectively 1, 2 and 5 components. At T_3 , we simulate from an independent sampler affecting 10 components. At T_4 , we simulate from a global independent sampler. In other words, proportionally to the temperature, we propose larger moves, using global samplers at the two highest temperatures. Simulating modifying less components at low temperature allows to have better acceptance rates in our high dimensional space ($M = 50$) and to allow local exploration of the posterior. Moreover, the

moves at the highest temperatures affect more components to improve the mixing of these chains, feeding the chains $(X^{(0)})$, $(X^{(1)})$, $(X^{(2)})$ with states, that they could not have attained without the exchange steps.

Regarding the proposal rate of information exchange, there is again a balance to do between high and low rate. A high rate will encourage information exchange, but will slow local exploration. Conversely, a low rate will hamper the process of exchanging information.

After a few experiments, we allowed the chains to exchange information according to the PIR scheme just after the first iteration with a probability of 0.1.

We ran this algorithm for 10000 iterations and we now describe its numerical results. The algorithm took 2 days to run on a desktop computer with a single processor AMD Opteron 146 2.0GHz.

5.3 Numerical results

We first show in figure 19 the energy of the states generated by the 5 chains used in the PIR for 10000 iterations.

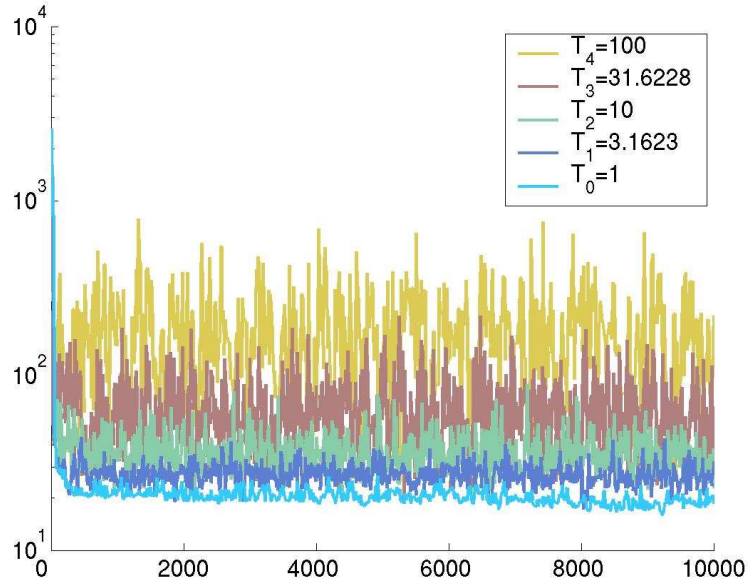


Fig. 19 Energy of the states of the 5 chains, generated by the PIR along the 10000 iterations.

Figure 19 shows the energy of the states of the 5 chains in function of the number of the iteration. we look at the lower curve, corresponding to T_0 , we can see that it stabilizes after about 500 iterations, around levels of energy corresponding to the expected order of magnitude of the mean energy of the posterior. Indeed, allowing exchanges since the beginning helps to converge quickly. As all the other chains show a stabilized profile of energy after this number of iterations, we consider it as the end of the burn-in period, that is, we consider that each chain is in stationary regime beyond this number of iterations. Moreover, we can see that each couple of chains at adjacent temperatures show overlapping energy profiles, allowing the exchanges between the two chains. Indeed, the empirical exchange acceptance rate has been found between 0.6 and 0.8 for each couple of adjacent chains.

We present in figure 20 the water cut curves corresponding to the states generated by the chain at T_0 after the 500th iteration. We plot the curves corresponding to this chain only, because the weights (21) of the states generated by the other chains are negligible in this application. We also represent the reference water cut as the thick line in black on this figure. In figure 21, the reference water cut in black, the median of the sample in red and the 95% percentile interval as red dot lines are represented. We use the median as estimator of the expectation as it is more robust than the empirical mean, particularly to eventual extreme values.

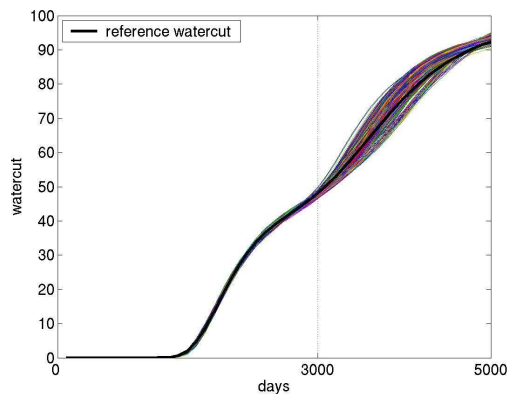


Fig. 20 Water Cut curves of the sample and reference.

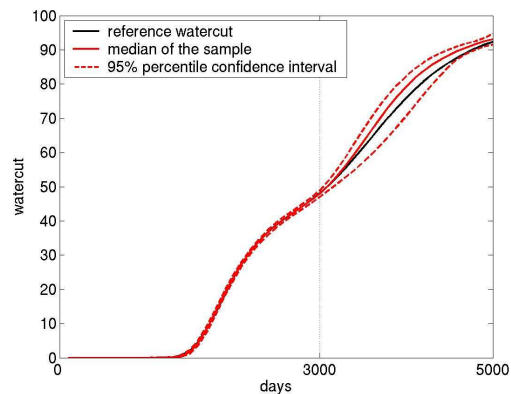


Fig. 21 Median, 95% percentile confidence interval and reference water cut.

We can see in figure 20 that the curves generated by our algorithm are well distributed around the reference one. In other words, the water cut data, up to 3000 days, is correctly matched for the sample. Particularly, the breakthrough time, that is the time when the water cut become greater than 0, is perfectly matched by the whole sample. Between 3000 and 5000, the curves are more widely spread around the reference, exploring the different possible prolongations of the curve. Comparing these results with figure 18 enlightens the quality of our results. In figure 21, we can also see that for the period matched (up to 3000 days), the median of the sample produced perfectly match the reference. Moreover, the 95% confidence interval is extremely thin around the reference water cut until 3000 days, when it widens according to the sample generated. We can see that for the next 2000 days, the reference water cut stays in the 95% confidence interval and is quite close to the median, validating our sample for prediction purposes. We can also remark that the confidence interval tends to shrink at the end of the period considered. This effect is due to the intrinsic nature of the water cut on this example. Indeed, it is a strictly increasing function with an asymptote at 100. This explains why the curves of the sample generated are getting closer at the end of the time period.

Finally, we represent in figure 22 the reference permeability field computed with all the components and a collection of eight realizations of the permeability field conditioned to the water cut data D^* , generated by the PIR algorithm.

We can first see in figure 22, that the aspect of the realizations (b., ..., i.) generated by our algorithm is far smoother than the reference. This is due to the approximation made by using a truncated KL expansion with $M = 50$ components. It has been seen in paragraph 3.4 that such an approximation is able to reproduce the dynamical properties of the field. It is then possible to add the remaining components of the KL expansion to each realization, simulating the remaining (ξ_i) , to obtain a permeability map which will have the same aspect as the reference one.

It can also be seen that the realizations we represent here are clearly different between each other. This illustrate the good exploration of the posterior (23) carried out by the PIR algorithm, by improving the mixing properties with respect to classical single MCMC. A single Markov chain could not have produced such various maps within the same number of iterations, see e.g. [7]. This ability is of great interest when addressing the problem of prediction uncertainties. Indeed, different permeability maps will lead to different prediction curves. Identifying numerous different permeability maps that are potential representations of the true permeability field will lead to better uncertainty assessment. That is illustrated here with the water cut curve prediction with its associated confidence interval on figure 21.

6 Conclusions

In this work, we have first described a way to reduce the dimension of the inverse problem in history matching by using a truncated KL expansion of the random field of interest. We have validated the method by a monte-carlo study. Then we have shown the deficiencies of classical MCMC methods when applied to multi-modal ill-posed inverse problems. We have then proposed an innovative application of recent stochastic simulation methods, based on parallel interacting Markov Chains. Numerical results on a toy example have shown its interest in the resolution of inverse problem in the Bayesian framework. Finally, an application on a synthetic

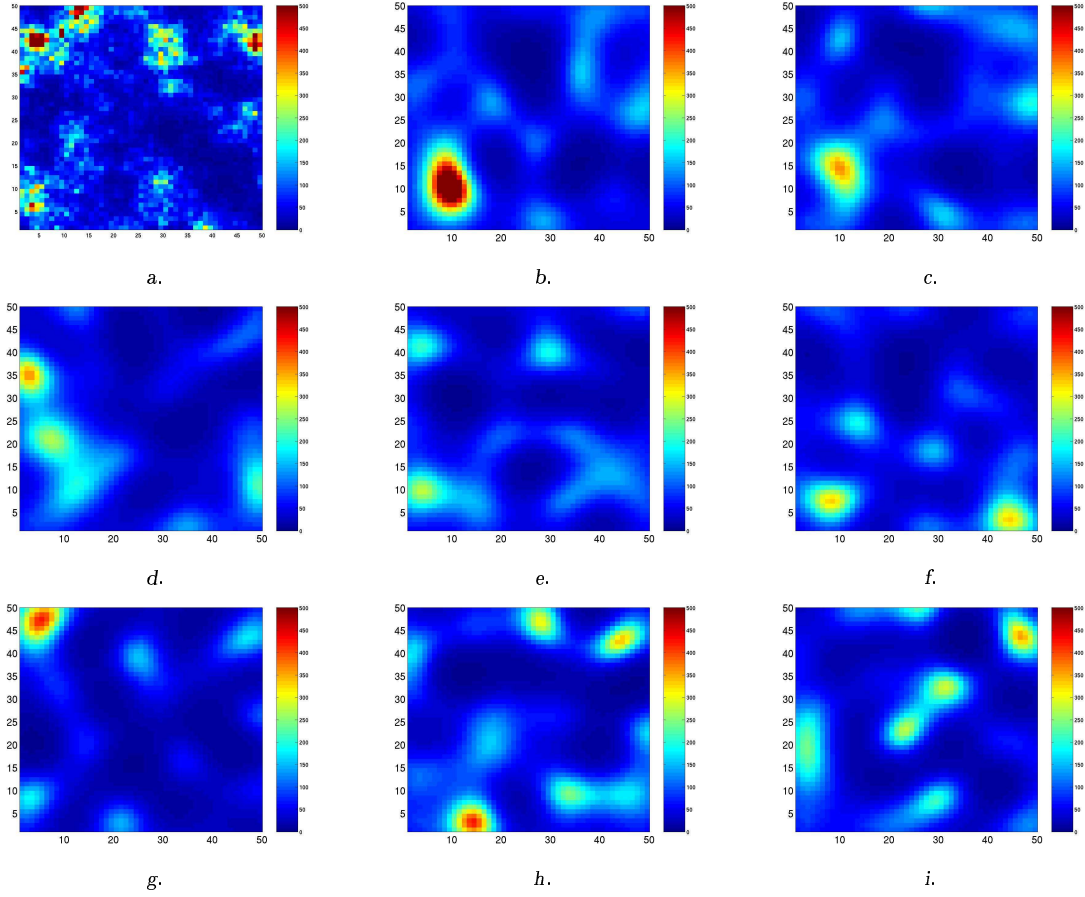


Fig. 22 Reference permeability field (a.) and 8 realizations from the posterior generated by the PIR algorithm (b. to i.).

case of reservoir characterization has been performed, and its results are clearly satisfactory. Especially, the results have demonstrated the improved properties of the PIR algorithm in the sampling of the posterior. Moreover, we have produced with this algorithm a sample with good predictive properties. Finally, we want to recall here that our method has the same computational cost as classical MCMC methods.

Improvements can be made on the parameterization of the parallel algorithm. For instance, it could be really interesting to use a Langevin sampler at the lowest temperature, as this kernel is known to exhibit the best local performances among the different samplers. However, it requires to compute the gradient of the fluid-flow simulator. This could be done by the use of a simulator with adjoint states or numerically. The dimension being still 50 in the case studied here, a numerical estimate of the gradient in each direction is intractable: it would require 50 additional simulations at each iteration to compute the local gradient estimate in single precision. Nevertheless, choosing the most important directions in which to compute the gradient should be of great interest. A preliminary sensitivity analysis could indeed be performed before running the algorithm, in order to evaluate the influence of the components of the KL expansion for the considered response of the fluid-flow simulator. Then, a local numerical estimate of the gradient could be computed in the directions chosen. This task is also easily parallelizable. With this estimation, the Langevin sampler could be implemented without increasing the computational cost. Theoretically it would make the local exploration of the posterior more accurate.

When new data are available, we have to integrate them. This problem of integrating new data could be easily addressed in the following way: we could use either the same method using the kernel given by the final estimation of (20) or an importance sampling resampling scheme ([24]) proposing a realization with the weights given by (21), then reweighting them according to the new results.

It would be also of great interest to investigate the case of Gaussian based lithofacies geostatistical models, like truncated Gaussian model or pluri-Gaussian model. As Gaussian related, a KL expansion could be

performed on these models in order to reduce their dimensionality. Then the sampling techniques addressed here could be used to perform the history matching.

Acknowledgements The author is very grateful to J. Jacod and E. Moulines for helpful discussions. Comments made by L.Y. Hu and D. Busby were also very helpful.

References

1. Andrieu, C., Jasra, A., Doucet, A., Moral, P.D.: On non-linear Markov Chain Monte Carlo via self-interacting approximations. Tech. rep., University of Bristol (2007)
2. Andrieu, C., Moulines, E.: On the Ergodicity Properties of some Adaptive MCMC Algorithms. *Annals of applied Probability* **16**, 1462–1505 (2003)
3. Atchadé, Y.: A cautionary tale on the efficiency of some adaptive monte carlo schemes. preprint (2007)
4. C. Andrieu and C. Robert: Controlled MCMC for optimal sampling. Tech. rep., Céremade, Université de PARIS - DAUPHINE (2001)
5. Caers, J., Hoffman, T.: The probability perturbation method: a new look at bayesian inverse modeling. *Mathematical Geology* **38**(1), 81–100 (2006)
6. Chen, Y., Zhang, D.: Data assimilation for transient flow in geologic formations via ensemble Kalman filter. *Advances in Water Resources* **29**, 1107–1122 (2006)
7. Dostert, P., Efendiev, Y., Hou, T., Luo, W.: Coarse-gradient Langevin algorithms for dynamic data integration and uncertainty quantification. *Journal of computational physics* **217**(1), 123 – 142 (2006)
8. Earl, D., Deem, M.: Parallel Tempering : Theory, Applications, and New Perspectives. *Physical Chemistry Chemical Physics* **7**, 3910 – 3916 (2005)
9. Gavalas, G., Shah, P., Seinfeld, J.: Reservoir history matching by Bayesian estimation. *SPE Journal* **16**(6), 337–350 (1976)
10. Geyer, C.: Markov Chain Monte Carlo Maximum Likelihood. In: *Computing Science and Statistics: Proceedings of 23rd Symposium on the Interface Interface Foundation*, p. 156. American Statistical Association, Fairfax Station, New-York (1991)
11. Ghanem, R., Spanos, P.: *Stochastic Finite Elements, a Spectral Approach*. Springer-Verlag (1991)
12. Haario, H., Saksman, E., Tamminen, J.: An Adaptive Metropolis Algorithm. *Bernoulli* **7**, 223–242 (2001)
13. Holden, L., Sannan, S., Soleng, H., Arntzen, O.: History matching using adaptive chains. Tech. rep., Norwegian Computing Center (2002)
14. Hu, L.Y.: Gradual Deformation and Iterative Calibration of Gaussian-Related Stochastic Models. *Mathematical Geology* **32**(1), 87–108 (2000)
15. Iba, Y.: Extended Ensemble Monte Carlo. *International Journal of Modern Physics C* **12**(5), 623 – 656 (2001)
16. Kitaniidis, P.: Quasi-linear geostatistical theory for inverting. *Water resources research* **31**(10), 2411–2419 (1995)
17. Kou, S., Zhou, Q., Wong, W.: Equi-energy sampler with applications in statistical inference and statistical mechanics. *The Annals of Statistics* **34**(4), 1581–1619 (2006)
18. Lehoucq, R., Sorensen, D., Yang, C.: Arpack user's guide: solution of large scale eigenvalue problems with implicitly restarted Arnoldi method (1997)
19. Loève, M.: *Probability Theory*. Princeton University Press (1955)
20. Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A.T.M.: Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21**, 1087–1091 (1953)
21. Oliver, D.: On conditional simulation to inaccurate data. *Mathematical Geology* **28**(6), 811–817 (1996)
22. Oliver, D., Reynolds, A., Bi, Z., Abacioglu, Y.: Integration of production data into reservoir models. *Petroleum Geoscience* **7**(9), 65–73 (2001)
23. Oliver, D.S., Cunha, L.B., Reynolds, A.C.: Markov Chain Monte Carlo Methods for Conditionning a Permeability Field to Pressure Data. *Mathematical Geology* **29**(1), 61–91 (1997)
24. Robert, C., Casella, G.: *Monte-Carlo Statistical Methods*, 2nd edn. Springer (2004)
25. Romary, T., Hu, L.: Assessing the dimensionality of random fields with Karhunen-Loève expansion. In: *Petroleum Geostatistics 2007*. EAGE (2007)
26. Sarma, P., Durlofsky, L., Aziz, K., Chen, W.: Efficient Real-time Reservoir Management Using Adjoint-based Optimal Control and Model Updating. *Computational Geosciences* **10**(1), 3–36 (2006)
27. StreamSim Technologies, Inc.: *3dsl User Manual*, Version 2.10 edn. (2003)
28. Tarantola, A.: *Inverse Problem Theory and Model Parameter Estimation*. SIAM (2005)

4.2 History matching of truncated Gaussian models
by parallel interacting Markov chains on a reduced
dimensional space

**History matching of truncated Gaussian models by parallel interacting Markov chains
on a reduced dimensional space**
Thomas Romary, Lin-Ying Hu

Abstract

In oil industry and subsurface hydrology, geostatistical models are often used to represent the spatial distribution of different lithofacies in the reservoir. Two main model families exist: multipoint and truncated Gaussian models. We focus here on the latter.

In history matching of lithofacies reservoir model, we attempt to find multiple realizations of lithofacies configuration, that are conditional to dynamic data and representative of the model uncertainty space. This problem can be formalized in the Bayesian framework. Given a truncated Gaussian model as a prior and the dynamic data with its associated measurement error, we want to sample from the conditional distribution of the facies given the data.

A relevant way to generate conditioned realizations is to use Markov Chains Monte Carlo (MCMC). However, the dimension of the model and the computational cost of each iteration are two important pitfalls for the use of MCMC. In practice, we have to stop the chain far before it has scanned the whole support of the posterior. Further more, as the relationship between the data and the random field is non-linear, the posterior can be multimodal. Hence, the chain may stay stuck in one of the modes.

In this work, we first show how to reduce drastically the dimension of the problem by using a truncated Karhunen-Loève expansion of the Gaussian random field underlying the lithofacies realization. Then we show how we can improve the mixing properties of classical single MCMC, without increasing the global computational cost, by the use of parallel interacting Markov chains at different temperatures. Applying the dimension reduction and this innovative sampling method lowers drastically the number of iterations needed to sample efficiently from the posterior. We show the encouraging results obtained when applying the methodology to a synthetic history matching case.

Introduction

Conditioning the spatial distribution of different lithofacies in the reservoir to production data, such as cumulative oil production, water cut, is a highly challenging task in reservoir modeling. It consists in solving an ill-posed inverse problem: given a prior knowledge on the random field governing the lithofacies spatial distribution in the reservoir, typically a geostatistical model, we aim at finding multiple realizations of this model that will exhibit the same dynamical behaviour of the true reservoir. In other words, we want to sample from the posterior distribution defined in the Bayesian inversion framework, see [17]. The dynamical behaviour of a given realization is computed by a fluid-flow simulator F .

The Markov chain Monte-Carlo (MCMC) are known to produce samples of virtually any posterior distribution. They have already a long story in history matching, see [14] for a review. Although MCMC methods seem particularly suited for our problem, there are nevertheless three obstacles. First, the dimension of the problem is generally the number of grid blocks on which is discretized the random field, often several hundreds of thousands. Therefore, the chain has to be run for an intractable number of iterations to converge and to achieve an efficient sampling of the posterior. Moreover, as the forward problem is highly non-linear, the chain can be trapped into a local mode, or even not achieve convergence. Finally, the computational cost of a single fluid flow simulation limits the affordable number of iterations.

Other methods have been proposed in the literature for the history matching of lithofacies models, see *e.g.* [8], [4], [12]. However, they result in history-matched realizations that are not necessarily representative of the posterior, whereas we propose here a method to sample exhaustively from the posterior.

In this work, we first propose a way to reduce the dimension of the problem using the Karhunen-Loève (KL) expansion ([13]) of geostatistical models. Second, we propose a method to improve the global efficiency of the MCMC by generating a collection of chains in parallel at different temperatures and allowing them to interact. We then apply this method to a synthetic history matching case and compare the results to those given by classical MCMC methods.

Dimension reduction

In this section, we describe the application of KL expansion to Gaussian random fields and subsequently to thresholded Gaussian models. We denote here the truncated Gaussian model by thresholded Gaussian model to avoid confusion with the truncated KL representation.

Let $X \in \mathcal{X}$ be a random process, indexed on $U \subset \mathbb{R}^d$, the physical domain of the reservoir, with mean μ and covariance Γ . The KL's theorem [13] proposes to represent X in the basis of its covariance's eigenfunctions (ϕ_i) . Then we have the following representations:

- i) $\xi_i = \int_U X_u \phi_i(u) du$,
- ii) $X_u = \sum_{i \in I} \xi_i \phi_i(u)$,
- iii) $\text{var}(\xi_i) = \lambda_i$, where λ_i is the eigenvalue associated with ϕ_i ,
- iv) If X_u is a Gaussian process, the (ξ_i) are independent Gaussian random variables with variance λ_i .

Truncating the sum in ii), we can construct approximations:

$$X_u^{(M)} = \sum_{i=1}^M \xi_i \phi_i(u) \quad (1)$$

of the true process that minimize the integrated mean squared error on \mathbf{U} : $\varepsilon_M^2 = \int_U E(X_u^{(M)} - X_u)^2 du$, as can be seen in [7]. Indeed, the λ_i being in decreasing order, the first related terms will represent the major part of the global variance. It is worth noticing that the components

(ξ_i) are independent if and only if X is Gaussian. Otherwise, although the (ξ_i) are uncorrelated, they are not independent, and their dependence links need to be studied carefully before using the KL expansion.

To reduce the dimension of thresholded Gaussian models, we apply the decomposition to the underlying Gaussian random field. In other words, denoting $Y = t(X)$ the facies random field, with t being the threshold function and X a Gaussian random field, we will consider $Y^{(M)} = t(X^{(M)})$, with $X^{(M)}$ given in (1). Furthermore, this method can be extended easily to pluri-Gaussian models (see *e.g.* [11]): considering the model $Y = t(X_1, X_2)$ where X_1 and X_2 are two Gaussian random fields, eventually correlated, we will use the following approximation: $Y^{(M)} = t(X_1^{(M)}, X_2^{(M)})$.

As shown in [16], we can determine the necessary number of components to reproduce a certain proportion of the variance of the process. For simulation of lithofacies however, it does not seem to be a good criterion. Other criteria such as reproduction of connexions between facies should be considered. In general, a preliminary study should be performed before choosing the number of components to use.

Parallel interacting Markov chains

The principle of making interact Markov chains first appears in [6] under the name parallel tempering (PT). The principle of PT is to simulate a number $(K + 1)$ of replica of the system of interest by MCMC, each at a different temperature in the sense of the simulated annealing, and to allow the chains to exchange information, swapping their current state. The high temperature systems are generally able to sample large volumes of state space, whereas low temperature systems, whilst having precise sampling in a local region of the state space, may become trapped in local energy minima during the timescale of a typical computer simulation. Parallel tempering achieves good sampling by allowing the systems at different temperatures to exchange their state. Thus, the inclusion of higher temperature systems ensures that the lower temperature systems can access a representative set of low-temperature regions of the state space. It is also worth noticing that PT can make efficient use of large CPU clusters, where different replicas can be run in parallel, unlike classical MCMC sampling that are sequential methods. An additional benefit of the PT method is the generation of results for a range of temperatures, which may also be of interest to the investigator. It is now widely appreciated that PT is a useful and powerful computational method.

More recently, some researchers from the statistical community have been investigating PT and more generally interacting Markov chains. They propose a general theoretical framework and new algorithms in order to improve the above exchange information step. Two main algorithms drawn our attention: the equi-energy sampler (EES) of [10] and the population importance-resampling MCMC sampler (PIR) of [2] so called in [3], which allows to go back in the history of the chain. More precisely, these two last algorithms are based on self interacting approximations of non-linear Markov kernels, defined by Andrieu et al. in [2]. We will focus here on the PIR algorithm, as it exhibits the most appealing properties for our application domain.

We first define the family $\{\pi^{(l)}, l = 0..K\}$ of distributions we want to simulate from, such that:

$$\pi^{(l)}(x) \propto e^{-E_l(x)}, \quad (2)$$

where $E_l(x) = \frac{E(x)}{T_l}$, $E(x) = \frac{1}{2}\|D^* - F(x)\|_{C_D^{-1}}^2 + \frac{1}{2}\|x - \mu\|_{\Gamma^{-1}}^2$ is the energy of the system and T_l is the temperature at which it is considered, $x \in \mathcal{X}$, D^* denotes the dynamical data, with its associated error covariance matrix C_D , and F denote the fluid flow simulator. The T_l satisfy: $T_0 = 1 < T_1 < \dots < T_K < +\infty$, so that $\pi^{(0)} = P(X|D^*)$, the posterior distribution of our inverse problem. These distributions are a family of *tempered* versions of $P(X|D^*)$. We

will also call E_l *tempered energies*. The PIR algorithm aims to simulate from:

$$\Pi(x) = \prod_{l=0}^K \pi^{(l)}(x^{(l)}), \quad (3)$$

where $x = (x^{(0)}, \dots, x^{(K)}) \in \mathcal{X}^{K+1}$, allowing exchanges between states at different temperatures. The dilated versions of $\pi^{(0)}$: $\pi^{(1)}, \dots, \pi^{(K)}$ are easier to simulate and can provide information on $\pi^{(0)}$.

We denote by $x = (x^{(0)}, \dots, x^{(K)}) \in \mathcal{X}^{K+1}$ the current state of the chain. For $l = 0, \dots, K-1$, we define the importance function as follows:

$$r^{(l)}(x^{(l)}) = e^{-(E_l(x^{(l)}) - E_{l+1}(x^{(l)}))} \quad (4)$$

The method can be formalized by defining the following kernel P_n at time n , given all the previous states $x_{0:n-1} = (x_0, \dots, x_{n-1})$ and for $A_0 \times \dots \times A_K \in \mathcal{B}(\mathcal{X}^{K+1})$:

$$P_n(x_{0:n-1}; A_0 \times \dots \times A_K) = P^{(K)}(x^{(K)}, A_K) \prod_{l=0}^{K-1} P_{x_{0:n-1}}^{(l+1)}(x^{(l)}, A_l), \quad (5)$$

where for $x_{0:n-1}^{(l+1)} \in \mathcal{X}^n$, $x^{(l)} \in \mathcal{X}$ and $A \in \mathcal{B}(\mathcal{X})$:

$$P_{x_{0:n-1}}^{(l+1)}(x^{(l)}; A) = \theta P^{(l)}(x^{(l)}, A) + (1 - \theta) \int_{\mathcal{X}} \nu_{x_{0:n-1}}^{(l+1)}(x^{(l)}, dy) T^{(l)}(y, x^{(l)}; A), \quad (6)$$

and

$$\nu_{x_{0:n-1}}^{(l+1)}(x^{(l)}, dy) = \frac{\sum_{i=0}^{n-1} r^{(l)}(x_i^{(l+1)}) \delta_{x_i^{(l+1)}}(dy)}{\sum_{i=0}^{n-1} r^{(l)}(x_i^{(l+1)})} \quad (7)$$

and in the algorithms considered here $T^{(l)}$ will take the following form:

$$T^{(l)}(y, x^{(l)}; A) = \min \left(1, \frac{r^{(l)}(y)}{r^{(l)}(x^{(l)})} \right) \mathbb{1}_A(y) + \left(1 - \min \left(1, \frac{r^{(l)}(y)}{r^{(l)}(x^{(l)})} \right) \right) \mathbb{1}_A(x^{(l)}) \quad (8)$$

In other words, at time step n , at temperature T_l , with probability θ , a classical Metropolis-Hastings (MH) move will be performed, and with probability $(1-\theta)$, an exchange move will be proposed. It consists in choosing a state y among $x_{0:n-1}^{(l+1)}$ with the weights $r^{(l)}(x_i^{(l+1)})$. This move is then accepted with probability $\min \left(1, \frac{r^{(l)}(y)}{r^{(l)}(x^{(l)})} \right)$. This step is an importance re-sampling step: we simulate from $\pi^{(l)}$, choosing a state simulated from the instrumental distribution $\pi^{(l+1)}$ (a state generated by the chain X^{l+1}), according to the importance weights $r^{(l)}(x_i^{(l+1)})$.

The main idea behind the PIR algorithm is that the kernel defined in (6) will converge towards the following limiting kernel:

$$P_{x_{0:n-1}}^{(l+1)}(x^{(l)}; A) = \theta P^{(l)}(x^{(l)}, A) + (1 - \theta) R^{(l)}(x^{(l)}, A), \quad (9)$$

where $R^{(l)}$ is a MH kernel, whose proposal distribution is given by:

$$Q^{(l)}(x^{(l)}, dy) = \pi^{(l)}(dy) \quad (10)$$

Obviously, the convergence towards $R^{(l)}$ will not be achieved in the time of the simulation, but its approximation at time n will help to sample from the posterior, particularly to span a larger part of the state space. Moreover, an exchange step will induce a *regeneration* of the chain that

will improve the mixing properties. In other words, the pieces of chain between two exchange steps will be independent one from the other, reducing the autocorrelations along the chain.

Finally, it is worth noticing that for this algorithm, we can use the samples of all the chains, reweighting them by the following importance weights:

$$\eta^{(l)}(x^{(l)}) = e^{-(E_0(x^{(l)}) - E_l(x^{(l)}))}, \quad (11)$$

to compute estimates of $I_g = E_{\pi_0}[g(X)]$, for any function g . Hence, the estimate \hat{I}_g , after N iteration of the algorithm will take the form:

$$\hat{I}_g = \sum_{l=0}^K \frac{\sum_{i=0}^N \eta^{(l)}(x_i^{(l)}) g(x_i^{(l)})}{\sum_{i=0}^N \eta^{(l)}(x_i^{(l)})}. \quad (12)$$

Concerning the choice of the parameters, although practical rules cannot be found for the PIR algorithm in the literature, some heuristic rules exist and are discussed in *e.g.* [9] for the PT algorithm and in [10] for the EES. However, the choice depends mainly on the problem addressed and there is no general recipe to tune the parameters. We will explain how we choose them for the history matching application in the next section.

Application to history matching

We consider the case where the prior on the lithofacies distribution is a 2-dimensional facies thresholded Gaussian model, with the following characteristics:

- its size is $2500 \times 2500 m^2$,
- it is discretized on a regular grid of $N = 50 \times 50$ blocks,
- it is 10 m thick,
- the underlying Gaussian random field has an isotropic covariance structure with a range equal to 600 m (a quarter the field edge size)
- it is composed of two lithofacies: A (50% with permeability 500md) and B (50% with permeability 10md),
- we put two wells in this field: an injector at grid node $(3, 3)$ and a producer at $(48, 48)$,
- the porosity is assumed constant at 0.25.

The field is assumed to be saturated in oil at time 0. The fluid flow is simulated with *3DSL* [1] during 5000 days with an injection rate at 5000 m^3/day and a pressure of 200 *bars* at the producer.

Given a reference realization of the field X^* , computed with all its 2500 components, and its water cut D^* calculated on 2000 days, we attempt to condition the geostatistical model X to the water cut D^* . According to the results of a preliminary study on the necessary number of components to reproduce the dynamical behaviour of such a model (see [16]), we choose to use a truncated KL expansion (1) with $M = 100$ components to represent the field. Hence, the dimension of the inference problem is reduced from 2500 to 100. The posterior takes the following form:

$$P(X^{(M)} | D^*) \propto e^{\left(-\frac{1}{2} \|D^* - F(X^{(M)})\|^2 - \frac{1}{2} \|X^{(M)} - \mu\|_{\Gamma_{(M)}^{-1}}^2 \right)}, \quad (13)$$

where $\Gamma_{(M)} = \Phi_{(M)} \Lambda \Phi_{(M)}^t$,
 $\Phi_{(M)}$ is the matrix $L \times M$ whose column vectors are the $\phi_i(x)$,
 Λ is the diagonal matrix $M \times M$ whose diagonal components are the λ_i .

Here, D^* and $F(X^{(M)})$ are both functions of time. The covariance of the measurement error on the water cut is assumed to be the identity matrix. We represent the reference realization of the field considered here in figure 1. We also represent in figure 2 the reference water cut curve together with a sample of 200 curves computed for a sample of 200 independent realizations of the prior.

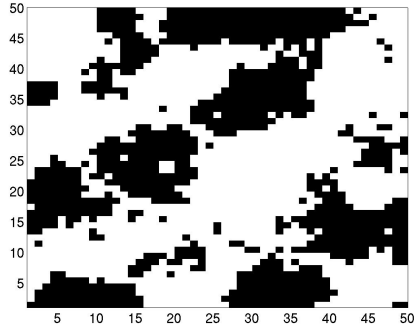


Figure 1: Reference lithofacies map.

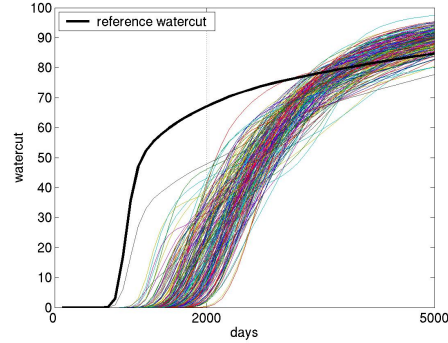


Figure 2: Water cut curves of an i.i.d. sample and reference.

We can see in figure 1 that there is an important portion of highly permeable (500 md) facies (in white) in the diagonal axis linking the two wells. Figure 2 shows its particular water cut profile: after the early water breakthrough, the water cut increases very fast, then slows down. This profile is very different from that of the curves of 200 independent realizations of X . Indeed, the minimum energy calculated for this sample is about 3000, with an average around 20000, whereas we expect energies around 50 for the matched sample. That makes our problem very challenging. In order to solve it, that is to sample from (13), we implement the PIR algorithm and a classical componentwise independent MH algorithm, that we will call single chain (SC) algorithm. The choice of the different parameters of the PIR algorithm is inspired by practical considerations found in [9] and [10] and a preliminary study of the energies of an i.i.d. sample of X .

We use 5 different temperatures, distributed geometrically between $T_0 = 1$ and $T_4 = 400$. A geometric distribution of the temperatures is a classical choice in the parallel tempering literature, see *e.g.* [5]. If the number of chains is sufficient, it allows generally a good overlapping of the histograms of the tempered energies, inducing the occurrence of exchanges. Namely, we take $T_l = T_0 \left(\frac{T_4}{T_0} \right)^{l/4}$ for $l = 1, 2, 3$. Hence, we use the the following temperature ladder:

$$T_0 = 1.000 < T_1 = 4.729 < T_2 = 22.361 < T_3 = 105.737 < T_4 = 400.000$$

Thus, we simulate the 5 Markov chains $(X^{(l)})$ at the temperature T^l . At T_0 , T_1 , T_2 , we simulate from a symmetric increments random walk MH algorithm with a step variance $0.15\sqrt{T_l}$, affecting respectively 5, 20 and 50 components. At T_3 , we simulate from an independent sampler affecting 80 components. At T_4 , we simulate from a global independent sampler. In other words, proportionally to the temperature, we propose larger moves, using global samplers at the two highest temperatures. Modifying less components at low temperature results in better acceptance rates in our high dimensional space ($M = 100$) and allows local exploration of the posterior. Moreover, the moves at the highest temperatures affect more components, thus improve the mixing of these chains and feed the chains $(X^{(0)})$, $(X^{(1)})$, $(X^{(2)})$ with states, that they could not have attained without the exchange steps. After a few experiments, we allowed the chains to exchange information according to the PIR scheme just after the first iteration with a probability of 0.05.

We ran both algorithm for 10000 iterations. The PIR algorithm took 50 hours to run on a desktop computer with a single processor AMD Opteron 146 2.0GHz and the SC algorithm took about 10 hours. Note that having implemented the PIR algorithm on a parallel computer architecture, it would have taken the same time as the SC.

In figure 3 a and b, we represent respectively the energy of the states of the 5 chains used in the PIR algorithm, and the energy of the states generated by the single chain.

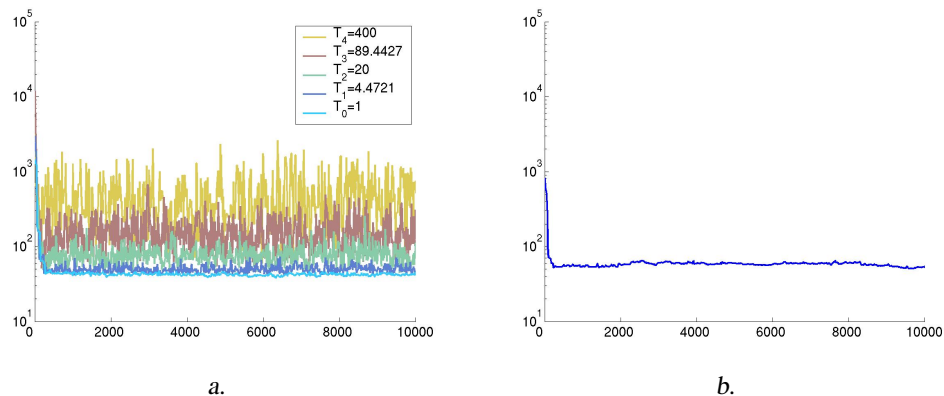


Figure 3: Energy of the states as a function of the number of iterations (a: 5 chains by PIR; b: SC).

Figure 3 a shows the energy of the states of the 5 chains as a function of the number of iteration. For the lower curve, corresponding to T_0 , we observe a stabilization after about 200 iterations, around levels of energy corresponding to the expected order of magnitude of the posterior mean energy. Indeed, allowing exchanges since the beginning of the chain accelerates its convergence. As all the other chains show a stabilized profile of energy after this number of iterations, we consider it as the end of the burn-in period, namely each chain is in stationary regime beyond this number of iterations. Moreover, we can see that each couple of chains at adjacent temperatures show overlapping energy profiles, allowing the exchanges between the two chains. Indeed, the empirical exchange acceptance rate has been found between 0.6 and 0.8 for each couple of adjacent chains.

Figure 3 b shows that the SC algorithm exhibits a rather fast convergence towards the stationary regime, attaining energy levels around 50 in about 250 iterations.

Figure 4 a and b, show some statistics computed from the samples generated respectively by the PIR algorithm and by the SC algorithm, namely, the median and the 95% percentile confidence interval of the water cut curves generated together with the reference water cut.

In figure 4 a, we can see that for the matched period (up to 2000 days), the median of the sample water cut perfectly matches the reference. Moreover, the 95% confidence interval is extremely thin around the reference water cut until 2000 days. Then it widens for the next 3000 days. In addition, the reference water cut stays in the 95% confidence interval and is quite close to the median. This validates our sample for prediction purposes.

Conversely, in figure 4 b., although the reference water cut is also correctly matched by the sample generated by the SC algorithm, its prediction abilities are rather weak with respect to the PIR algorithm: the confidence interval generated is still thin beyond 2000 days and does not include the reference water cut. This is due to the only local exploration performed by this algorithm.

Figure 5 shows 7 realizations by the PIR and one by SC. First, the aspect of the realizations is far smoother than the reference. This is due to the approximation by a truncated Karhunen-Loève expansion with $M = 100$ components. Second, the realizations generated by PIR (a to g) are clearly different between each other (we did not reproduce here the whole variety of

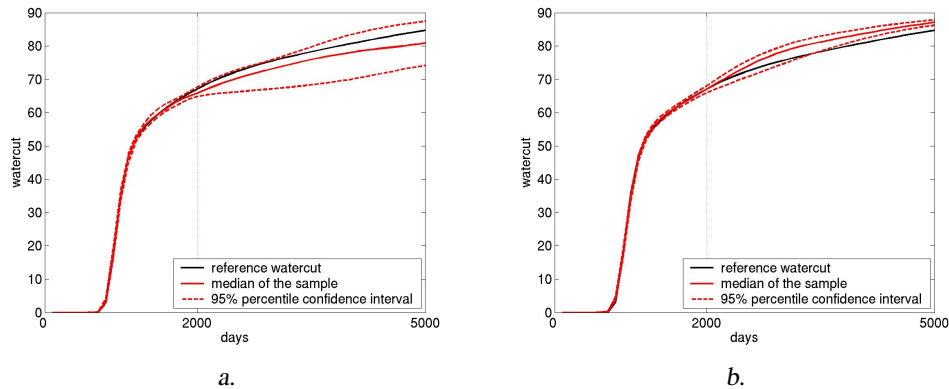


Figure 4: Median, 95% percentile confidence interval and reference water cut (a. PIR algorithm, b. SC).

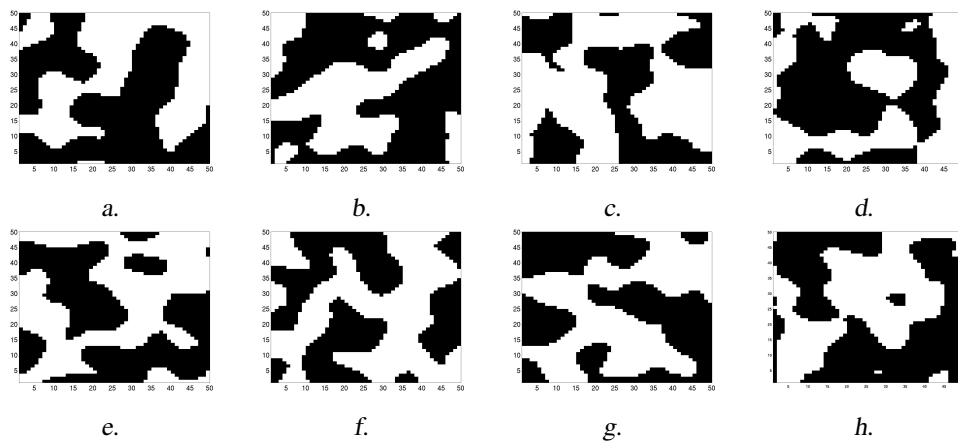


Figure 5: 7 realizations from the posterior generated by the PIR (a. to g.) and one generated by the SC (h.).

maps generated). This illustrates the good exploration of the posterior (13) carried out by the PIR, due to the improved mixing properties with respect to classical single MCMC. Finally, all realizations generated by the SC are similar between each other (figure 5 h). Note that all the maps generated by both algorithms reproduce a link of highly permeable facies between the two wells.

Besides, it is worth noticing that the PIR exhibits a global empirical acceptance rate of about 0.4, whereas the SC shows a empirical acceptance rate around 0.1. In other words, comparatively, we throw away twice more fluid-flow simulations with the SC than with the PIR.

Conclusion

In this work, we have first described a way to reduce the dimension of the inverse problem in history matching of thresholded Gaussian models by using a truncated KL expansion of the underlying Gaussian random field. We have then proposed an innovative application of a recent stochastic simulation method, based on parallel interacting Markov chains. Finally, an application on a synthetic case of reservoir modeling has been performed. The numerical results show the advantage of the PIR algorithm for sampling the posterior, with respect to a single Markov chain. The sampling carried out by PIR explores better the posterior, therefore the sample produced has a better capacity of prediction. Moreover, this method is well suited for parallel computing, thus comparable with the classical MCMC in terms of computation costs.

Further improvements can be made on the parameterization of the parallel algorithm. It should be of great interest to imagine an automatic tuning of the kernels parameters, namely the number of components affected at each iteration and the variance step used in random walks, according to the temperature.

Finally, the problem of integrating new data in an existing model can be performed in the following way: we could use either the above method with the kernel given by the final estimation of (9) or an importance sampling resampling scheme ([15]). The latter consists in proposing a realization with the weights given by (11), then reweighting them according to the adequation to new data.

References

- [1] [2008]. *3dsl User Manual*. StreamSim Technologies, Inc., Version 3.00 edition.
- [2] Andrieu, C., Jasra, A., Doucet, A., and Moral, P. D. [2007]. On non-linear Markov chain Monte Carlo via self-interacting approximations. Technical report, University of Bristol.
- [3] Atchadé, Y. [2007]. A Cautionary Tale on the Efficiency of some Adaptive Monte Carlo Schemes. *preprint*.
- [4] Caers, J. [2003]. History Matching under Training-Image Based Geological Model Constraints. *SPE Journal*, 8(3):218–226.
- [5] Earl, D. and Deem, M. [2005]. Parallel Tempering : Theory, Applications, and New Perspectives. *Physical Chemistry Chemical Physics*, 7:3910 – 3916.
- [6] Geyer, C. [1991]. Markov Chain Monte Carlo Maximum Likelihood. In *Computing Science and Statistics: Proceedings of 23rd Symposium on the Interface Interface Foundation*, page 156. American Statistical Association, Fairfax Station, New-York.
- [7] Ghanem, R. and Spanos, P. [1991]. *Stochastic Finite Elements, a Spectral Approach*. Springer-Verlag.
- [8] Hu, L. Y. [2000]. Gradual Deformation and Iterative Calibration of Gaussian-Related Stochastic Models. *Mathematical Geology*, 32(1):87–108.
- [9] Iba, Y. [2001]. Extended Ensemble Monte Carlo. *International Journal of Modern Physics C*, 12(5):623 – 656.
- [10] Kou, S., Zhou, Q., and Wong, W. [2006]. Equi-Energy Sampler with Applications in Statistical Inference and Statistical Mechanics. *The Annals of Statistics*, 34(4):1581–1619.
- [11] Lantuéjoul, C. [2002]. *Geostatistical simulation*. Springer.
- [12] Liu, N. and Oliver, D. [2005]. Critical Evaluation of the Ensemble Kalman Filter on History Matching of Geologic Facies. *SPE Journal*, 8(6):470–477.
- [13] Loève, M. [1955]. *Probability Theory*. Princeton University Press.
- [14] Oliver, D., Reynolds, A., Bi, Z., and Abacioglu, Y. [2001]. Integration of Production Data into Reservoir Models. *Petroleum Geoscience*, 7(9):65–73.
- [15] Robert, C. and Casella, G. [2004]. *Monte-Carlo Statistical Methods*. Springer, 2nd edition.
- [16] Romary, T. and Hu, L. [2007]. Assessing the Dimensionality of Random Fields with Karhunen-Loève Expansion. In *Petroleum Geostatistics 2007*. EAGE.
- [17] Tarantola, A. [2005]. *Inverse Problem Theory and Model Parameter Estimation*. SIAM.

4.3 Reservoir forecasting under uncertainty

RESERVOIR FORECASTING UNDER UNCERTAINTY

D. Busby¹, T. Romary¹, S. Touzani¹, M. Feraille¹, B. Nøtinger¹, L. Hu¹

¹IFP (Institut Français du Pétrole), France;

daniel.busby@ifp.fr

Abstract

This paper proposes an efficient methodology to deal with uncertainty in Reservoir Forecasting applications. Different tools such as Gaussian process based on non-parametric response surface models, sequential experimental design, Karhunen-Loève (KL) decomposition of random fields and Global Sensitivity Analysis (GSA) are integrated in an efficient uncertainty workflow.

The proposed workflow should replace expensive and sometimes inefficient Monte Carlo methods, involving direct fluid flow simulator, traditionally used to handle uncertainty. GSA is, respect to the traditional local SA, a much more powerful tool that can be used to better understand the reservoir by recognizing and quantifying the main sources of uncertainty or for adjusting controllable parameters in Enhanced Oil Recovery applications.

Introduction

Output of multiphase fluid flow simulators used in reservoir forecasting, must be used with caution by practitioners because of uncertainty. Inputs of such simulators are derived from indirect inverse measurements or are given by expert reservoir engineers or geologists and their values can be highly uncertain. Propagating this uncertainty, from the input to the output of such simulators, is a very difficult task requiring massive computational power. Traditional Monte Carlo methods are often unaffordable and need to be replaced by more efficient statistical methods.

In more formal terms, the output of the simulator, is denoted by the vector $\mathbf{F}(\{\mathbf{x}\}, \{\mathbf{y}\})$, depending on the variables \mathbf{x} and \mathbf{y} . The set \mathbf{x} represents collectively the uncertain parameters describing the reservoir: permeability, porosity etc... In a geostatistical context, \mathbf{x} may be a very large dimensionality vector representing a random function of position, and is thus characterised by a multidimensional probability density. The set of parameters \mathbf{y} are man-controlled control parameters (e.g. number and position of wells, a choice of recovery scenario etc..).

The final goal is to recover the most oil at the lowest technical price with minimum risk and uncertainties. This means that one wants to estimate the average forecast $\langle \mathbf{F}(\{\mathbf{x}\}, \{\mathbf{y}\}) \rangle_{\{\mathbf{x}\}} = \langle \mathbf{F} \rangle_{\{\mathbf{x}\}}(\{\mathbf{y}\})$, where the average $\langle \dots \rangle_{\{\mathbf{x}\}}$ is taken with respect to the whole set of parameters $\{\mathbf{x}\}$ describing the reservoir. It remains a dependence on the exploitation parameters $\{\mathbf{y}\}$. Estimation of the forecast uncertainties $\langle \delta \mathbf{F}^2(\{\mathbf{x}\}, \{\mathbf{y}\}) \rangle_{\{\mathbf{x}\}} = \langle \mathbf{F}^2(\{\mathbf{x}\}, \{\mathbf{y}\}) \rangle_{\{\mathbf{x}\}} - \langle \mathbf{F} \rangle_{\{\mathbf{x}\}}^2(\{\mathbf{y}\})$ are also needed.

In a next step, the optimal exploitation scheme $\{\mathbf{y}\}^*$ is an acceptable solution of an optimisation problem: $\{\mathbf{y}\}^* = \arg \min \{J\{\mathbf{y}\}\}$, where $J\{\mathbf{y}\}$ is a prescribed oil company dependent function of the mean estimation, $\langle \mathbf{F} \rangle_{\{\mathbf{x}\}}(\{\mathbf{y}\})$, and of its uncertainty $\langle \delta \mathbf{F}^2(\{\mathbf{x}\}, \{\mathbf{y}\}) \rangle_{\{\mathbf{x}\}}$. The oil company strategy specifies $J\{\mathbf{y}\}$. History matching techniques in a Bayesian context using know reservoir dynamic data can be used to restrict the distribution of \mathbf{x} and the resulting uncertainties.

So far, due to the high dimensionality of the considered problems, and to the highly non linear dependence of \mathbf{F} , fulfilling this program is far beyond the most powerful computing facilities.

The statistical design and analysis of complex simulators output is used to sensibly reduce the number of required simulations by building a statistical approximation, often referred to as a response surface model or emulator, of the simulator output. To build such approximation, a number of runs are performed on a carefully selected input configuration (the experimental design) and the results are then analysed by mean of parametric or non-parametric regression techniques. This work focus on non-parametric regression techniques based on Gaussian process, that allows to approximate complex functions. A sequential design strategy can also be used to improve the emulator accuracy at minimum cost.

The emulator is then used to propagate uncertainty and for global Sensitivity Analysis (SA) of both adjustable and uncertain parameters. Global SA is used to decompose the output variance in terms of all the input uncertain factors (i.e. parameters and interaction between parameters). Knowledge of the relative contribution of an input can be useful to focus on the most important parameters to subsequently try to reduce their uncertainty.

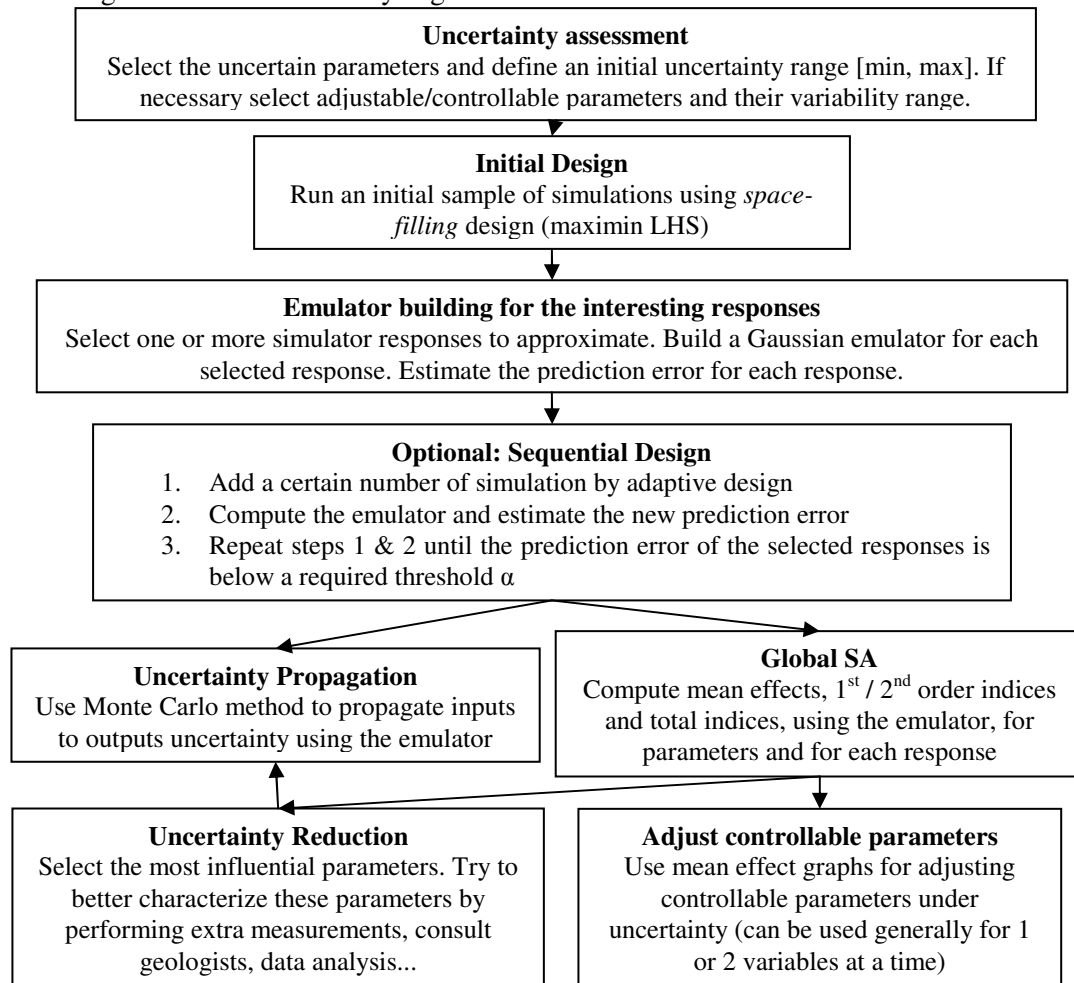
In previous works different type of emulators have been used to propagate uncertainty of continuous parameters [Busby et al. (1), Scheidt et al. (5), Zabalza et al. (6)]. In this work a model reduction approach based on the KL decomposition [Loève (11)] is proposed to also treat random field inputs such as permeability and porosity. The KL expansion is a parameterization of a random field which allows to effectively reduce its dimensionality. Numerical results show that only a small number of components of the truncated KL expansion can be used to effectively represent the major part of the variance of the field while preserving the spatial variability. The eigenvalues associated to this expansion are then the uncertain inputs of the simulator.

By performing a global sensitivity analysis using the Gaussian emulation approach it can be observed that only a few of the KL components are influential on the cumulative oil production forecasts.

1. Methodology Description

1.1. Overview

In this Section a schematic representation of the proposed uncertainty analysis workflow is presented. In this workflow only continuous parameters uncertainty is taken into account. An indication of how to treat geostatistical uncertainty is given in Section 1.4.



1.2. Gaussian emulator for response surface approximation

Given an unknown function $f(\cdot)$ and set of observations $f_X = \{f(\mathbf{x}_1), f(\mathbf{x}_2), \dots, f(\mathbf{x}_N)\}$ at design sites $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ the objective is to approximate the value of f at an unknown point \mathbf{x} . The method used in this approach is based on kriging or more generally on a best linear unbiased estimator discussed in a previous work [Busby et al. (1)]. We remind that in kriging the following model is adopted:

$$S(x) = \sum_{j=1}^m \beta_j h_j(x) + Z(x) \quad \text{eq.1}$$

where the regression part is a linear combination of pre-selected real-valued functions h_1, \dots, h_m , with unknown coefficients β_1, \dots, β_m . Z is assumed to be a centred Gaussian random process of covariance

$$\text{cov}[x, y] = E[Z(x)Z(y)] = \sigma^2 R(x, y),$$

where $\sigma^2 = E[Z(x)^2]$ denotes the process variance, and $R(x, y)$ is the correlation function. Note that as discussed in O'Hagan et al. (2), a Bayesian approach can also be adopted starting from the model defined at eq.1 by providing prior distributions of the parameters β and σ^2 , then a corresponding posterior distribution of the response $f(x)$ can be derived.

As discussed in previous works, we prefer to work with the following correlation function:

$$R(x, y) = \exp\left(-\sum_{j=1}^d \left(\frac{|x_j - y_j|}{\theta_j}\right)^p\right)$$

which is stationary. As discussed in Santner et al. (4) the parameters θ and p are determined by *maximum likelihood* estimation. The experimental design proposed in this approach is a maximin Latin Hypercube design where the number of simulations depends on the number of uncertain parameters and on the complexity of the function to analyze, but in practice is usually given by the available CPU time. After building the emulator using the available simulator runs, the prediction accuracy of the emulator can be estimated by leave-one-out cross validation. Note that if this accuracy is not sufficiently high a sequential design strategy such as the one discussed in Busby et al. (1) or in Scheidt et al. (5) can be adopted. In these approaches further simulations are generally added at local maxima of the kriging variance. Here, we restrict ourselves to the use of sequential design.

1.3. Global Sensitivity Analysis

Global Sensitivity Analysis (GSA) is the study of how the variability of the model input has an effect on the variability of the output. As discussed in Saltelli et al. (9) and O'Hagan et al. (8) GSA is based on the Sobol' decomposition of a function $f(x)$.

In order to describe this method, let us suppose that a mathematical model is described by a function $f(x)$, $x = (x_1, \dots, x_p)$, and is defined in a unit p -dimensional cube: $\Omega^p = \{x | 0 \leq x_i \leq 1; i = 1, \dots, p\}$

The main idea according to Sobol's [Sobol (17)] approach is to decompose $f(x_1, \dots, x_p)$ into summands of different dimensions:

$$f(x_1, \dots, x_p) = f_0 + \sum_{i=1}^p f_i(x_i) + \sum_{1 \leq i < j \leq p} f_{ij}(x_i, x_j) + \dots + f_{1,2,\dots,p}(x_1, \dots, x_p) \quad (2)$$

where f_0 is a constant, and the integrals of every summand over any of its own variables must be zero,

$$\text{i.e. } \int_0^1 f_{i_1, \dots, i_s}(x_{i_1}, \dots, x_{i_s}) dx_{i_k} = 0, \text{ where } 1 \leq i_1 < \dots < i_s \leq p, \quad s = 1, \dots, p \text{ and } 1 \leq k \leq s.$$

It follows from this definition that

$$f_0 = \int_{\Omega^p} f(x) dx$$

and all the summands in (2) are orthogonal, i.e if $(i_1, \dots, i_s) \neq (j_1, \dots, j_l)$, then

$$\int_{\Omega^p} f_{i1,...,is} f_{j1,...,jl} dx = 0 \quad (3)$$

Sobol showed that the decomposition of $f(x_1, \dots, x_p)$ is unique and that all the terms in equation (2) can be evaluated via multidimensional integrals:

$$f_i(x_i) = -f_0 + \int_{\Omega^{p-1}} f(x) dx^i$$

$$f_{i,j}(x_i, x_j) = -f_0 - f_i(x_i) - f_j(x_j) + \int_{\Omega^{p-2}} f(x) dx^{ij}$$

where dx^i and dx^{ij} are the product $dx_1 \dots dx_p$ without dx_i and $dx_i dx_j$, respectively. Same kind of formulae can be obtained for the higher-order terms.

By squaring and integrating equation (2) over Ω^p , and by using equation (3), we obtain the total variance of $f(x)$ as a summation of the variance associated to each summands f_i :

$$V = \sum_{i=1}^p V_i + \sum_{1 \leq i < j \leq p} V_{ij} + \dots + V_{1,2,\dots,p} \quad (4)$$

The total variance V of $f(x)$ is also defined as $V = \int_{\Omega^p} f^2(x) dx - f_0^2$

Then, to explain the part of output variance due to parameters or interactions between parameters the following sensitivity indices can be built:

$$S_{i1,...,is} = \frac{V_{i1,...,is}}{V} \text{ for } 1 \leq i1 < \dots < is \leq p$$

S_i is called the first order sensitivity index for factor x_i , which measures the part of output variance explained by the effect of x_i .

$S_{i,j}$, for $i \neq j$, is called the second order sensitivity index which measures the part of output variance due to the interaction effect between x_i and x_j .

The total sensitivity index S_{Ti} for one particular parameter x_i , defined as the sum of all the sensitivity indices involving the parameters [Saltelli et al. (18)], can also be very useful to measure the part of output variance explained by all the effects in which x_i play a role.

$$S_{Ti} = \sum_{k \# i} S_k \text{ where } \#i \text{ represented all the } i1, \dots, is \text{ terms that include the index } i.$$

GSA enables to explain the variability of the output response in function of the input parameters through the definition of total and partial sensitivity indices. These indices can be estimated using Monte Carlo or Quasi Monte Carlo techniques to approximate the different multidimensional integrals, requiring large sampling. Therefore GSA techniques cannot be used directly using fluid flow simulator. In this work the computation of the required sensitivity coefficients are performed using emulators of interesting responses. These emulators are built as described in §1.2.

1.4. Karhunen-Loève expansion of random fields

In this part we describe a technique that enables to parameterize a realization of porosity or permeability using a few parameters. The parameterization is based on the Karhunen-Loève expansion applied to Gaussian random fields (GRF) which is mainly based on the following proposition:

Proposition

Let X_t , $t \in \mathbf{T}$, a zero mean second order real process with covariance Γ . There exists an Hilbert space \mathcal{H}_0 of real functions on \mathbf{T} such that the space \mathcal{H}_0 spanned by the functions $\Gamma(t, \cdot)$, $t \in \mathbf{T}$, is dense in \mathcal{H} .

This space is called the Reproducing Kernel Hilbert Space (RKHS) of X or of kernel Γ . It is isomorphic to the subspace \mathbf{H} of $L^2(\Omega, \mathbf{A}, \mathbf{P})$ generated by the random variables X_t . \mathcal{H} being a Hilbert space, it admits a countable orthonormal basis (φ_i) . A judicious choice of the base (φ_i) allows obtaining interesting properties. Particularly, the Karhunen-Loève's theorem [Loève (11)] proposes to use the family of eigenfunctions of Γ as an orthonormal family of \mathcal{H} . Indeed, Γ being positive

definite, Mercer's theorem on positive definite kernels ensures existence and uniqueness of this decomposition. The theorem states as follow:

Karhunen-Loève Theorem:

Let X_t , $t \in \mathbf{T}$ be a zero mean second order real process with covariance Γ . If Γ is a continuous covariance on $\mathbf{T} \times \mathbf{T}$, there exists an orthonormal basis of \mathcal{H} spanned by eigenfunctions (φ_i) of $\Gamma(t, \cdot)$. The corresponding basis of \mathbf{H} is the family (ξ_i) of zero mean, uncorrelated random variables. In particular, we have the following representations :

- i. $\xi_i(\omega) = \int_{\mathbf{T}} X_t(\omega) \varphi_i(t) dt$,
- ii. $X_t(\omega) = \sum_{i \in I} \xi_i(\omega) \varphi_i(t)$,
- iii. $\text{Var}(\xi_i) = \lambda_i$, where λ_i is the eigenvalue associated with $\varphi_i(t)$,
- iv. If X_t is a Gaussian process, the (ξ_i) are independent Gaussian random variable with variance λ_i .

Practically, truncating the sum in the ii and sorting the eigenvalues in descending order, we construct approximations of the true process: $\hat{X}_t^{(N)}(\omega) = \sum_{i=1}^N \xi_i(\omega) \varphi_i(t)$, that minimize the integrated mean squared error on \mathbf{T} : $\varepsilon_M^2 = \int_{\mathbf{T}} (X_t^N - X_t)^2 dt$, as can be seen in [Ghanem et Spanos (12)].

Indeed, the λ_i being in increasing order, the first related terms will represent the major part of the global variance of the GRF. We can see on Figure 1, which corresponds to application 2 described later, that using the first 30 components of the decomposition we are taking into account almost 80% of the global variance of the GRF.

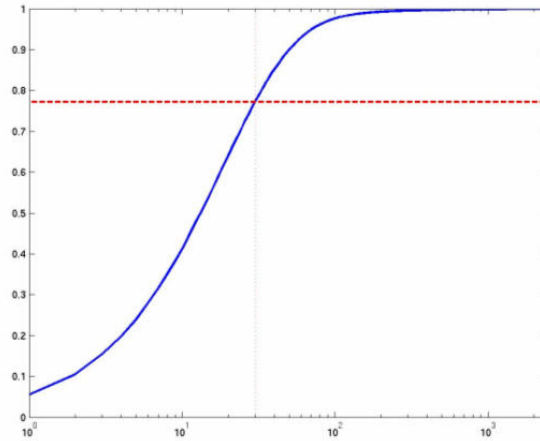


Figure 1: Gaussian process variance reconstruction as a function of the number of components

2. Applications to reservoir forecasting

In this Section some reservoir forecasting applications of the proposed approach are presented. The first application consists in testing the quality of our emulator on the response surface of a very non-linear synthetic reservoir test case: the IC Fault Model. The second application consists in determining on a 2D synthetic case which components of the KL permeability field decomposition are the more sensitive onto the production forecasts.

2.1.IC Fault Model

The IC Fault Model is a simple synthetic reservoir model presented in [Tavassoli et al. (14)].

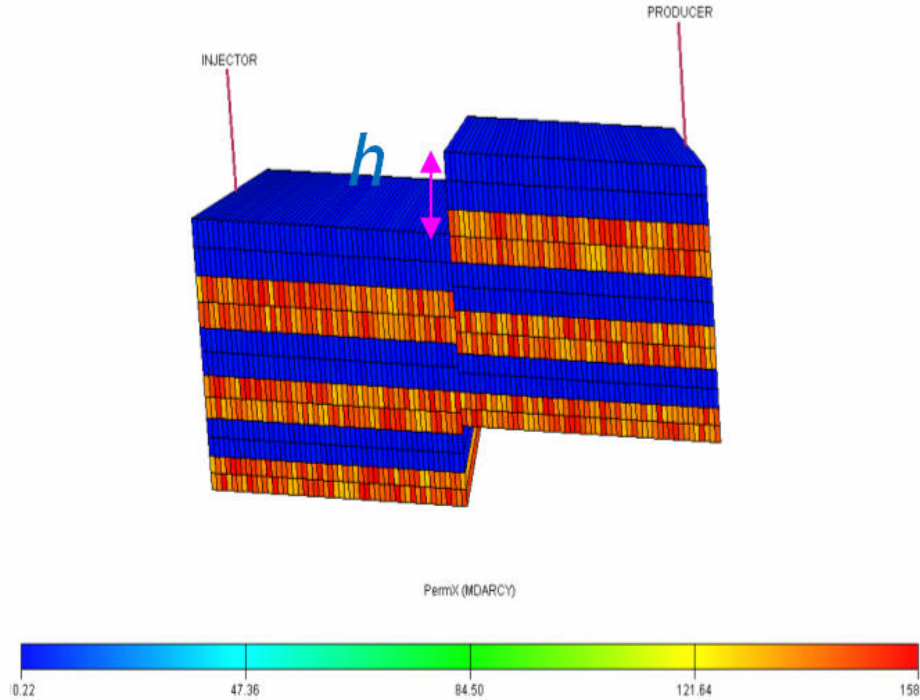


Figure 2: IC Fault model

In [Tavassoli et al. (14)] it was shown that even a simple model like this could be very difficult to history match. Irrespective of its apparent simplicity this model is in fact highly nonlinear and present different local minima. In the initial model problem [Tavassoli et al. (14)] only three uncertain parameters were considered:

- X1) h : fault throw [0,40] ft;
- X2) Kh : good sand horizontal permeability multiplier (red layers in the figure) [100, 200] md;
- X3) Kl : poor sand horizontal permeability multiplier (blue layers in the figure) [0, 50] md;

The following additional uncertain parameters are added to increase the complexity of this test case:

- X4) Ph : porosity of the good sand [0.12, 0.18];
- X5) Pl : porosity of the poor sand [0.24,0.36];
- X6) $KVKH$: vertical/horizontal permeability anisotropy (multiplier) [0.1,1];
- X7) Swc : connate water saturation [0.08,0.15];

We use our approximation scheme to approximate the response surface of the total oil production after 10 years (FOPT10). A total of 85 simulations obtained by a maximin LHS experimental design are used to build the emulator. The estimated mean squared prediction error obtained by cross-validation and using only the available simulations is equal to 1.75 %.

To check the accuracy of this result we performed 2000 random confirmation runs and compare the simulator results with the emulator ones. A scatter plot showing this result is given in Figure 3 below.

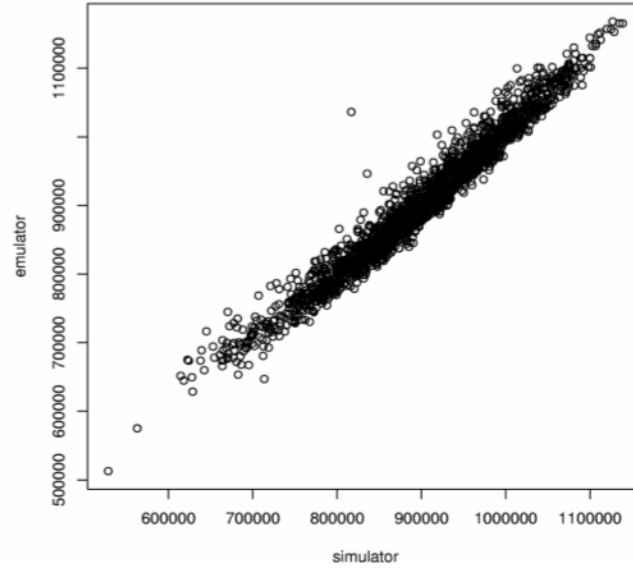


Figure 3: Scatterplot (2000 simulations) emulator-simulator for IC Fault model and FOPT10 response

Note also that the empirical root mean squared error η^2 : $\eta^2 = \sum_{i=1}^N \frac{(f(x_i) - S(x_i))^2}{N}$, where x_i , $i=1 \dots N$

are the confirmation runs f is the simulator output and S is the emulator, is equal to 2,3% and is therefore very close to our estimated one. A global sensitivity analysis is then performed using the emulator. First order sensitivity indices and the total effects are shown in Figure 4.

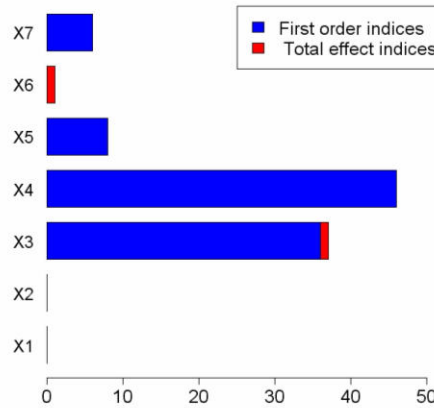


Figure 4: First order and total sensitivity indices for IC Fault model and FOPT10 response

Note that for this test case the total effects are very close to the first order effects meaning that there are no influential interactions of parameters for this response. The parameters accounting for 80% of the response variance are the poor layers permeability K1 and the good layers porosity Ph. Therefore a better characterization of these parameters is essential to reduce the total oil production uncertainty in this model. Moreover this results were not evident a priori before performing sensitivity analysis.

2.2. Geostatistical Uncertainty

In this section we analyze the uncertainty of a permeability random field using the KL parameterization discussed in Section 1.4 for a synthetic 2D reservoir model. The permeability field is a classical log-normal random field whose size is 2500 x 2500 m² with mean 44.7 md and

covariance given by: $\Gamma(t, s) = \sigma^2 \exp\left(-\left(\frac{\|t-s\|}{a}\right)^{1.9}\right)$, with a so as to have a range equal to a quarter the

length of the field and $\sigma = 0.5$, discretized on a regular grid of 50x50 cells. There are two wells in the

model one injector at location (3,3) and one producer at (48,48). The permeability values are conditioned to the wells with a value of 90 md. An example of one possible realization of the permeability field is shown on Figure 5.

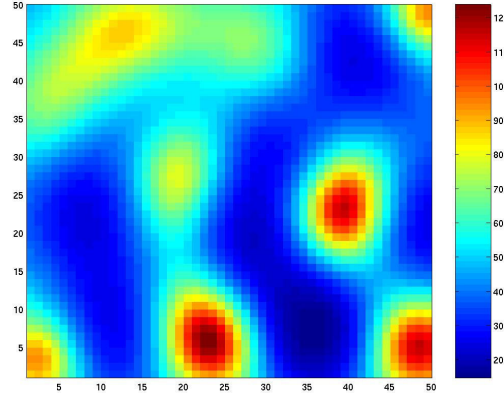


Figure 5: Example of one possible realization of the permeability field

The porosity is assumed constant 0.25. The field is assumed to be saturated in oil at time 0. The flow is simulated with 3DSL [13] during 10000 days with an injection rate at 5000 m³/day and a pressure of 200 bars at the producer.

As shown on previous Figure 1, in section 1.4, by truncating the KL expansion after the first 30 components we are able to reconstruct 80% of the permeability field variance. Our objective is to study the influence of each of the first 30 KL components on the total oil production after 10 years (FOPT10). To this end, a global sensitivity analysis is performed using the methodology discussed in sections 1.2 and 1.3 using the independent normally distributed variables ξ of the expansion as uncertain parameters. An experimental design of 200 simulations is selected by maximin LHS. An emulator is built for the response FOPT10 of interest, with estimated accuracy of 1.1%. This result indicates that the Gaussian emulator can reproduce the response very accurately in using only 200 simulations for 30 dimensions. To check the reliability of our accuracy estimate we made an additional 4000 confirmation runs (Figure 6) and compute the empirical root mean squared error η^2 as computed previously in section 2.1.

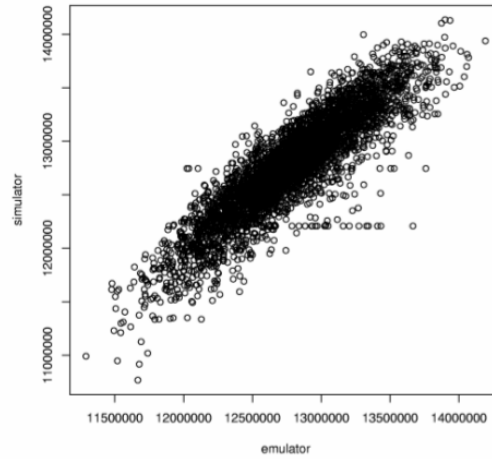


Figure 6: Scatterplot (4000 simulations) emulator-simulator for FOPT10 response

Note that we found an η^2 corresponding to 1.7% the FOPT10 response confirming our estimated accuracy. This checking confirms as well that 200 runs are enough to catch the response shape of this synthetic case. Using our accurate emulator a global sensitivity analysis is performed on FOPT10. Results of the first order and total effects sensitivity indices are shown on Figure 7.

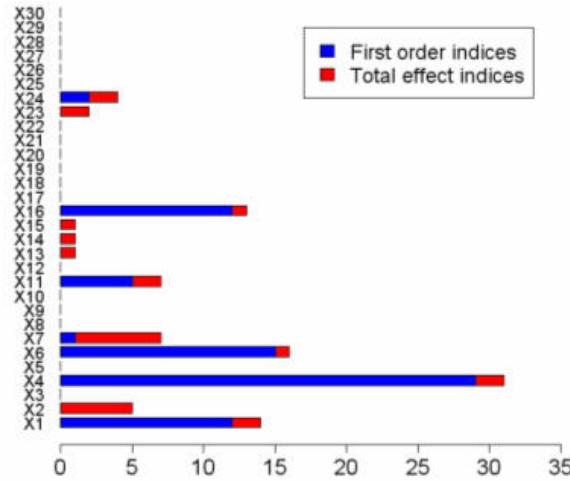


Figure 7: First order and total sensitivity indices for the 30 uncertain parameters, corresponding to the first components of the permeability field KL expansion, for FOPT10 response.

The results for the FOPT10 response show that only a few components are really influential onto the production forecasts 1, 2, 4, 6, 7, 11, 16, 24 these components are shown on Figure 8.

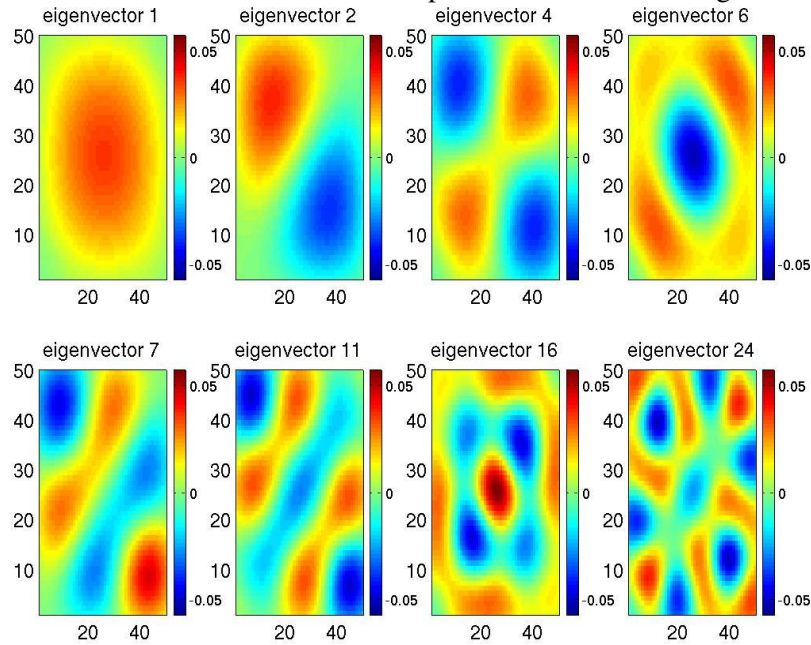


Figure 8: Eigenvectors associated to the eight most influential components of the KL expansion for FOPT10 response

The influential components on the oil production are linked with the production scheme, imposed rate and pressure as well as with the geometry of the reservoir. On our example all the components found as influential show a major role on the diagonal linking the injector with the producer.

Conclusions - Perspectives

In this paper, we have presented several methods organized in a sequential workflow allowing to quantify the influence of various parameters about reservoir forecasting. The idea is to select only the most relevant degrees of freedom that will control the forecasts. Variations may arise both from uncertainties in the reservoir description or from the control parameters values.

As the latter is the result of highly non linear processes that combine in a complex way, we must be able to control at each step the self-consistency of the retained assumptions. Some fundamental issues must be addressed:

- Detection of discontinuities: most standard tools of experimental design or GSA assume an underlying continuity. What happens in the discontinuous case?
A possible approach is to build a some kind of "phase diagram" of the problem allowing to detect the occurrence of discontinuities: percolation thresholds, marginal stability in case of multiphase flows.
- In a more risk oriented management, we could try to look for the possible occurrence of "catastrophic extreme scenario", e. g. to find the probability to have a recovery factor lower than any prescribed value.

Finally, in our quenched disorder context, is it possible, using relevant approaches (Zhang, 2002 and references therein), to avoid the costly average over the random permeability maps step? This task may be achieved using stochastic approaches that remain to be formalized for highly non linear transport equations.

References

1. Busby D., Farmer C.L., Iske A.: Hierarchical Nonlinear Approximation for Experimental Design and Statistical Data Fitting. *SIAM J. Sci. Comput.* 29, 1, 49-69, 2007.
2. Kennedy M., O'Hagan A.: Bayesian calibration of computer models (with discussion). *J. R. Statist. Soc. Ser. B Stat. Methodol.* 68, 425-464, 2001.
3. Sacks J, Welch W.J., Mitchell T.J., Wynn H.P.: Design and analysis of computer experiments (with discussion). *Statist. Sci* 4, 409-423, 1989.
4. Santner J., Williams B.J., Notz W.I.: *The Design and Analysis of Computer Experiments*. Springer Verlag 2003, New York.
5. Scheidt C., Zabalza-Mezghani I., Feraille M., Collombier D.: Adaptive Evolutive Experimental Designs for Uncertainty assessment - An innovative exploitation of geostatistical techniques ", IAMG, Toronto, 21-26 August, Canada, 2005.
6. Zabalza-Mezghani, E. Manceau, F. Roggero: "A New Approach for Quantifying the Impact of Geostatistical Uncertainty on Production Forecasts: The Joint Modeling Method", IAMG 2001 Conference, September 6-12, 2001.
7. E. Manceau and M. Mezghani and I. Zabablza-Mezghani and F. Roggero: "Combination of Experimental Design and Joint Modeling Methods for Quantifying the Risk Associated With Deterministic and Stochastic Uncertainties - An Integrated Test Study." Society of Petroleum Engineers, Paper N° 71620, 2001.
8. J. Oakley and A. O'Hagan: "Probabilistic sensitivity analysis of complex models: A Bayesian approach", *J. Roy. Statist. Soc. Ser.~B*, 16, pp.~751—769, 2004.
9. A. Saltelli, K. Chan and M. Scott: "Sensitivity Analysis", New York, Wiley, 2000.
10. P. Sarma, L.J. Durlofsky, K. Aziz, and W.H. Chen: "A new Approach to Automatic History Matching using Kernel PCA. In SPE Reservoir Simulation Symposium, 2007.
11. M.M. Loève: "Probability Theory". Princeton University Press, 1955.
12. R.G. Ghanem and P.D. Spanos: "Stochastic Finite Elements, a Spectral Approach", SV, 1991.
13. StreamSim Technologies, Inc.: "3dsl User Manual", Version 2.10, 2003.
14. Z. Tavassoli, J. N. Carter, and P. R. King: "Errors in history matching", *SPE Journal*, SPE paper n° 86883, 2004.
15. S. Subbey, M. Christie and M. Sambridge: "A strategy for Rapid Quantification of Uncertainty in Reservoir Performance Prediction", SPE paper n° 79678, 2004.
16. D. Zhang "Stochastic methods for Flow in Porous Media, Copying with Uncertainties" Academic Press 2002.
17. I.M Sobol: "Sensitivity estimates for nonlinear mathematical models". *Mathematical Modelling and Computational Experiments*, 1 :407-414, 1993.
18. A. Saltelli, S. Tarantola, F. Campolongo. "Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models". Wiley, 2004.

Bibliographie

- [3ds, 2003] (2003). *3dsl User Manual*. StreamSim Technologies, Inc., Version 2.10 edition.
- [Abramowitz and Stegun, 1965] Abramowitz, M. and Stegun, I. A. (1965). *Handbook of mathematical functions*. Dover Publications.
- [Adler and Taylor, 2006] Adler, R. and Taylor, J. (2006). *Random Fields and their Geometry*. To appear.
- [Alex et al., 2002] Alex, M., Vasilescu, O., and Terzopoulos, D. (2002). Multilinear Analysis of Image Ensemble : TensorFaces. In *European Conference of Computer Vision (ECCV '02)*.
- [Andrieu et al., 2007] Andrieu, C., Jasra, A., Doucet, A., and Moral, P. D. (2007). On non-linear Markov chain Monte Carlo via self-interacting approximations. Technical report, University of Bristol.
- [Andrieu and Moulines, 2003] Andrieu, C. and Moulines, E. (2003). On the Ergodicity Properties of some Adaptive MCMC Algorithms. *Annals of applied Probability*, 16 :1462–1505.
- [Armstrong et al., 2001] Armstrong, M., Galli, A., Loch, G. L., Geoffroy, F., and Eschard, R. (2001). *Pluri-Gaussian Simulations*. Kluwer, Dordrecht.
- [Atchadé, 2007] Atchadé, Y. (2007). A Cautionary Tale on the Efficiency of some Adaptive Monte Carlo Schemes. *preprint*.
- [C. Andrieu and C. Robert, 2001] C. Andrieu and C. Robert (2001). Controlled MCMC for optimal sampling. Technical report, Céremade, Université de PARIS - DAUPHINE.
- [Caers and Hoffman, 2006] Caers, J. and Hoffman, T. (2006). The probability perturbation method : a new look at bayesian inverse modeling. *Mathematical Geology*, 38(1) :81–100.
- [Celeux et al., 1999] Celeux, G., Hurn, M., and Robert, C. (1999). Computational and inferential difficulties with mixture posterior distributions. Technical Report RR-3627, INRIA.
- [Chavent et al., 1975] Chavent, G., Lemmonier, P., and Dupuy, M. (1975). History matching by use of optimal control theory. *SPE Journal*, 15(1) :74–86.
- [Chen et al., 2005] Chen, M., Zhang, D., Keller, A., and Lu, Z. (2005). A stochastic analysis of steady state two-phase flow in heterogeneous media. *Water resources research*, 41(1).

- [Chen et al., 2002a] Chen, Z., Micchelli, C., and Xu, Y. (2002a). Discrete wavelet Petrov-Galerkin methods. *Advances in Computational Mathematics*, 16 :1–28.
- [Chen et al., 2002b] Chen, Z., Micchelli, C., and Xu, Y. (2002b). Fast Collocation Methods For Second Kind Integral Equations. *SIAM J. NUMER. ANAL.*, 40(1) :344–375.
- [Chiles and Delfiner, 1999] Chiles, J.-P. and Delfiner, P. (1999). *Geostatistics, Modeling Spatial Uncertainty*. John Wiley & Sons.
- [De Lathauwer, 1997] De Lathauwer, L. (1997). *Signal Processing Based On Multilinear Algebra*. PhD thesis, Katholieke Universiteit Leuven.
- [Deutsch and Journel, 1994] Deutsch, C. and Journel, A. (1994). The Application of Simulated Annealing to Stochastic Reservoir Modeling. *SPE Advanced Technology Series*, 2(2) :222–227.
- [Dostert et al., 2006] Dostert, P., Efendiev, Y., Hou, T., and Luo, W. (2006). Coarse-gradient Langevin algorithms for dynamic data integration and uncertainty quantification. *Journal of computational physics*, 217(1) :123 – 142.
- [Earl and Deem, 2005] Earl, D. and Deem, M. (2005). Parallel Tempering : Theory, Applications, and New Perspectives. *Physical Chemistry Chemical Physics*, 7 :3910 – 3916.
- [Ecker, 2003] Ecker, M. (2003). *Encyclopedia of Life Support Systems (EOLSS)*, chapter Geostatistics : Past, Present and Future. Eolss Publishers, Oxford ,UK. Developed under the Auspices of the UNESCO.
- [Emery, 2004] Emery, X. (2004). *Simulation conditionnelle de modèles isofactoriels*. PhD thesis, Ecole des Mines de Paris.
- [Fang and Wang, 1994] Fang, K.-T. and Wang, Y. (1994). *Number-theoretic Methods in Statistics*. Chapman & Hall.
- [Gavalas et al., 1976] Gavalas, G., Shah, P., and Seinfeld, J. (1976). Reservoir History Matching by Bayesian Estimation. *SPE Journal*, 16(6) :337–350.
- [Genest and Remillard, 2004] Genest, C. and Remillard, B. (2004). Tests of independence and randomness based on empirical copula process. *Test*, 13(2) :335–369.
- [Geyer, 1991] Geyer, C. (1991). Markov Chain Monte Carlo Maximum Likelihood. In *Computing Science and Statistics : Proceedings of 23rd Symposium on the Interface Interface Foundation*, page 156. American Statistical Association, Fairfax Station, New-York.
- [Geyer and Thompson, 1995] Geyer, C. and Thompson, E. (1995). Annealing Markov Chain Monte Carlo with Applications to Ancestral Inference. *Journal of the American Statistical Association*, 90(431) :909–920.
- [Ghanem and Spanos, 1991] Ghanem, R. and Spanos, P. (1991). *Stochastic Finite Elements, a Spectral Approach*. Springer-Verlag.
- [Gunning, 2002] Gunning, J. (2002). On the Use of Multivariate Lévy-Stable Random Field Models for Geological Heterogeneity. *Mathematical Geology*, 34(1) :34–62.
- [Haario et al., 2004] Haario, H., Laine, M., Lethinen, M., Saksman, E., and Tamminen, J. (2004). Markov Chain Monte Carlo Methods for high dimensional inversion in remote sensing. *Journal of the Royal Statistical Society B*, 66(3) :1–17.

- [Haario et al., 2001] Haario, H., Saksman, E., and Tamminen, J. (2001). An Adaptive Metropolis Algorithm. *Bernoulli*, 7 :223–242.
- [Haario et al., 2003] Haario, H., Saksman, E., and Tamminen, J. (2003). Component-wise adaptation for MCMC. *Preprint*.
- [Handcock and Stein, 1993] Handcock, M. and Stein, M. (1993). A Bayesian Analysis of Kriging. *Technometrics*, 89(426) :368–390.
- [Holden, 1998] Holden, L. (1998). Adaptive chains. Technical report, Norwegian Computing Center.
- [Holden et al., 2002] Holden, L., Sannan, S., Soleng, H., and Arntzen, O. (2002). History matching using adaptive chains. Technical report, Norwegian Computing Center.
- [Hotelling and Pabst, 1936] Hotelling, H. and Pabst, M. R. (1936). Rank Correlation and Tests of Significance Involving no Assumption of Normality. *The Annals of Mathematical Statistics*, 7 :29–43.
- [Hu, 2000] Hu, L. Y. (2000). Gradual Deformation and Iterative Calibration of Gaussian-Related Stochastic Models. *Mathematical Geology*, 32(1) :87–108.
- [Hu and HERESIM, 1995] Hu, L. Y. and HERESIM, G. (1995). Vers un HERESIM non-stationnaire en 3d. Technical Report IFP : 41 943, IFP.
- [Huang et al., 2001] Huang, S., Quek, S., and Phoon, K. (2001). Convergence study of the truncated Karhunen-Loève expansion. *International journal for numerical methods in engineering*, 52 :1029–1043.
- [Hyvärinen and Oja, 2000] Hyvärinen, A. and Oja, E. (2000). Independent Component Analysis : Algorithms and Applications. *Neural Networks*, 13(4-5) :411–430.
- [Iba, 2001] Iba, Y. (2001). Extended Ensemble Monte Carlo. *International Journal of Modern Physics C*, 12(5) :623 – 656.
- [Jacod and Joathon, 1971] Jacod, J. and Joathon, P. (1971). Use of Random-Genetic Models in the study of Sedimentary Process. *Mathematical Geology*, 3(3) :219–233.
- [Kirby and Sirovich, 1995] Kirby, M. and Sirovich, L. (1995). Application of the Karhunen-Loève Procedure for the Characterization of Human Faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1) :103–108.
- [Kirkpatrick et al., 1983] Kirkpatrick, S., Jr., C. G., and Vecchi, M. (1983). Optimization by Simulated Annealing. *Science*, 220(4598) :671–680.
- [Kitanidis, 1995] Kitanidis, P. (1995). Quasi-linear geostatistical theory for inversing. *Water resources research*, 31(10) :2411–2419.
- [Kou et al., 2006] Kou, S., Zhou, Q., and Wong, W. (2006). Equi-Energy Sampler with Applications in Statistical Inference and Statistical Mechanics. *The Annals of Statistics*, 34(4) :1581–1619.
- [Le Ravalec et al., 2000] Le Ravalec, M., Noetinger, B., and Hu, L. (2000). The FFT Moving Average (FFT-MA) Generator : An Efficient Numerical Method for Generating and Conditioning Gaussian Simulations. *Mathematical Geology*, 32(6) :701–723.
- [Lopez, 2003] Lopez, S. (2003). *Modélisation de réservoirs chenalisés méandriiformes : approches génétiques et stochastiques*. PhD thesis, Ecole Nationale Supérieure des Mines de Paris.

- [Loève, 1955] Loève, M. (1955). *Probability Theory*. Princeton University Press.
- [Marinari and Parisi, 1992] Marinari, E. and Parisi, G. (1992). Simulated tempering : A new monte-carlo scheme. *Europhysics Letter*, 19 :451–458.
- [Marsily et al., 1984] Marsily, G. D., Lavedan, G., Boucher, M., and Fasanino, G. (1984). Interpretation of Interference Tests in a well field using Geostatistical Techniques to fit the Permeability Distribution in a Reservoir Model. In Verly, G., editor, *Geostatistics for Natural Resources Characterization, Part. 2*, pages 831–849. D. Reidel.
- [Matheron, 1965] Matheron, G. (1965). *Les variables régionalisées et leur estimation. Une application de la théorie des fonctions aléatoires aux Sciences de la nature*. Masson.
- [Matérn, 1986] Matérn, B. (1986). *Spatial Variation*. Springer-Verlag.
- [Mercer, 1909] Mercer, J. (1909). Functions of Positive and Negative Type, and their Connection with the Theory of Integral Equations. *Philosophical Transactions of the Royal Society of London. Series A*, 209 :415–446.
- [Metropolis et al., 1953] Metropolis, N., Rosenbluth, A., Rosenbluth, M., and Teller, A. T. M. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21 :1087–1091.
- [Nøtinger et al., 2005] Nøtinger, B., Artus, V., and Zargar, G. (2005). The future of stochastic and upscaling methods in hydrogeology. *Hydrogeology Journal*, 13 :184–201.
- [Oliver, 1996] Oliver, D. (1996). On conditional simulation to inaccurate data. *Mathematical Geology*, 28(6) :811–817.
- [Oliver et al., 2001] Oliver, D., Reynolds, A., Bi, Z., and Abacioglu, Y. (2001). Integration of Production Data into Reservoir Models. *Petroleum Geoscience*, 7(9) :65–73.
- [Oliver et al., 1997] Oliver, D. S., Cunha, L. B., and Reynolds, A. C. (1997). Markov Chain Monte Carlo Methods for Conditionning a Permeability Field to Pressure Data. *Mathematical Geology*, 29(1) :61–91.
- [Phoon et al., 2004] Phoon, K., Huang, S., and Quek, S. (2004). Comparison between Karhunen-Loeve and wavelet expansions for simulation of Gaussian processes. *Computers and structures*, 82 :985–991.
- [Press et al., 1992] Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. (1992). *Numerical Recipes in C : the Art of Scientific Computing*. Cambridge University Press.
- [Pycke, 2003] Pycke, J.-R. (2003). *Un lien entre le développement de Karhunen-Loève de certains processus gaussiens et le laplacien dans des espaces de Riemann*. PhD thesis, Université Paris 6.
- [Revuz, 1997] Revuz, D. (1997). *Probabilités*. Hermann.
- [Rice, 1945] Rice, S. (1945). Mathematical Analysis of Random Noise-Conclusion. *Bell Systems Tech. J.*, 24 :46–156.
- [Robbins and Monro, 1951] Robbins, H. and Monro, S. (1951). A Stochastic Approximation Method. *Annals of Mathematical Statistics*, 22(3) :400–407.

- [Robert and Casella, 2004] Robert, C. and Casella, G. (2004). *Monte-Carlo Statistical Methods*. Springer, 2nd edition.
- [Roberts and Rosenthal, 1998] Roberts, G. and Rosenthal, J. (1998). Optimal Scaling of Discrete Approximations to Langevin Diffusions. *Journal of the Royal Statistical Society B*, 60(1) :255–268.
- [Roberts and Rosenthal, 2006] Roberts, G. and Rosenthal, J. (2006). Examples of Adaptive MCMC. *Preprint*.
- [Roislien and Omre, 2006] Roislien, J. and Omre, H. (2006). T-distributed Random Fields : A Parametric Model for Heavy-tailed Well-log Data. *Mathematical Geology*, 38(7) :821–850.
- [Ruiz-Molina et al., 1999] Ruiz-Molina, J., Navarroa, J., and Valderrama, M. (1999). Differentiation of the modified approximative Karhunen-Loève expansion of a stochastic process. *Statistics & Probability Letters*, 42 :91–98.
- [Sarma et al., 2006] Sarma, P., Durlofsky, L., Aziz, K., and Chen, W. (2006). Efficient Real-time Reservoir Management Using Adjoint-based Optimal Control and Model Updating. *Computational Geosciences*, 10(1) :3–36.
- [Sarma et al., 2007] Sarma, P., Durlofsky, L., Aziz, K., and Chen, W. (2007). A new Approach to Automatic History Matching using Kernel PCA. In *2007 SPE Reservoir Simulation Symposium*.
- [Schneider and Willsky., 2003] Schneider, M. and Willsky., A. (2003). A Krylov Subspace Method for Covariance Approximation and Simulation of Random Processes and Fields. *Multidimensional Systems and Signal Processing*, 14(4) :195–318.
- [Spearman, 1904] Spearman, C. (1904). The Proof and Measurement of Association Between Two Things. *American Journal of Psychology*, 15(88).
- [Stramer and Tweedie, 1999] Stramer, O. and Tweedie, R. (1999). Self-Targeting Candidates for Metropolis-Hastings Algorithm. *Methodology and computing in Applied Probability*, 1(3) :307–328.
- [Strebelle, 2002] Strebelle, S. (2002). Conditional Simulation of Complex Geological Structures using Multiple-Point Geostatistics. *Mathematical Geology*, 34 :1–22.
- [Tarantola, 2005] Tarantola, A. (2005). *Inverse Problem Theory and Model Parameter Estimation*. SIAM.
- [Tierney, 1994] Tierney, L. (1994). Markov Chains for Exploring Posterior Distributions. *The Annals of Statistics*, 22(4) :1701–1762.
- [Tjelmeland et al., 1994] Tjelmeland, H., Omre, H., and Hegstad, B. (1994). Sampling from Bayesian models in Reservoir Characterization. Technical Report 2/1994, Norwegian Institute of Technology.
- [Watkins, 1993] Watkins, A. (1993). Linking Universal Kriging and Reservoir History Matching. In SPE, editor, *68th Annual Technical conference*, pages 9–20.
- [Zhang and Lu, 2004] Zhang, D. and Lu, Z. (2004). An efficient, high-order perturbation approach for flow in random porous media via Karhunen-Loève and polynomial expansions. *Journal of Computational Physics*, 194 :773–794.